POLS/CSSS 503
Advanced Quantitative Political Methodology

Models of Stationary & Non-Stationary Time Series

Christopher Adolph*
University of Washington, Seattle

May 19, 2011

*Assistant Professor, Department of Political Science and Center for Statistics and the Social Sciences.

The story so far

We've learned:

- why our LS models don't work well with time series
- the basics of time series dynamics

Next steps:

- Estimate AR(p), MA(q), and ARMA(p,q) models for stationary series
- Use our time series knowledge to select p and q
- Use simulations to understand how $\hat{y}_t$ changes as we vary $\mathbf{x}_t$

An AR(1) Regression Model

To create a regression model from the AR(1), we allow the mean of the process to shift by adding c_t to the equation:

$$y_t = y_{t-1}\phi_1 + c_t + \varepsilon_t$$

We then parameterize c_t as the sum of a set of time varying covariates,

$x_{1t}, x_{2t}, x_{3t}, \ldots$

and their associated parameters,

$\beta_1, \beta_2, \beta_3, \ldots$

which we compactly write in matrix notation as $c_t = \mathbf{x}_t\boldsymbol{\beta}$

An AR(1) Regression Model

Substituting for c_t, we obtain the AR(1) regression model:

$$y_t = y_{t-1}\phi_1 + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

Estimation is by maximum likelihood, *not* LS

(We will discuss the LS version later)

MLE accounts for dependence of y_t on past values; complex derivation

Let's focus on interpreting this model in practice

Interpreting AR(1) parameters

Suppose that a country's GDP follows this simple model

$$\text{GDP}_t = \phi_1 \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t$$

Interpreting AR(1) parameters

Suppose that a country's GDP follows this simple model

$$\begin{aligned} \text{GDP}_t &= \phi_1 \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\ \text{GDP}_t &= 0.9 \times \text{GDP}_{t-1} + 10 + 2 \times \text{Democracy}_t + \varepsilon_t \end{aligned}$$

Suppose that at year t, $\text{GDP}_t = 100$,
and the country is a non-democracy, $\text{Democracy}_t = 0$.

What would happen if we "made" this country a democracy in period $t + 1$?

Interpreting AR(1) parameters

$$y_t = y_{t-1}\phi_1 + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

Recall:
an AR(1) process can be viewed as the geometrically declining sum of all its past errors.

Interpreting AR(1) parameters

$$y_t = y_{t-1}\phi_1 + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

Recall:
an AR(1) process can be viewed as the geometrically declining sum of all its past errors.

When we add the time-varying mean $x_t\boldsymbol{\beta}$ to the equation, the following now holds:

$$y_t = (\mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t) + \phi_1(\mathbf{x}_{t-1}\boldsymbol{\beta} + \varepsilon_{t-1}) + \phi_1^2(\mathbf{x}_{t-2}\boldsymbol{\beta} + \varepsilon_{t-2}) + \phi_1^3(\mathbf{x}_{t-3}\boldsymbol{\beta} + \varepsilon_{t-3}) + \ldots$$

That is, y_t represents the sum of all past x_t's as filtered through $\boldsymbol{\beta}$ and ϕ_1

Interpreting AR(1) parameters

Take a step back: suppose c_t is actually fixed for all time at c, so that $c = c_t$

Interpreting AR(1) parameters

Take a step back: suppose c_t is actually fixed for all time at c, so that $c = c_t$

Now, we have

$$y_t = (c + \varepsilon_t) + \phi_1(c + \varepsilon_{t-1}) + \phi_1^2(c + \varepsilon_{t-2}) + \phi_1^3(c + \varepsilon_{t-3}) + \dots$$

Interpreting AR(1) parameters

Take a step back: suppose c_t is actually fixed for all time at c, so that $c = c_t$

Now, we have

$$\begin{aligned} y_t &= (c + \varepsilon_t) + \phi_1(c + \varepsilon_{t-1}) + \phi_1^2(c + \varepsilon_{t-2}) + \phi_1^3(c + \varepsilon_{t-3}) + \ldots \\ &= \frac{c}{1 - \phi_1} + \varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_1^2\varepsilon_{t-2} + \phi_1^3\varepsilon_{t-3} \ldots \end{aligned}$$

which follows from the limits for infinite series

Interpreting AR(1) parameters

Take a step back: suppose c_t is actually fixed for all time at c, so that $c = c_t$

Now, we have

$$\begin{aligned} y_t &= (c + \varepsilon_t) + \phi_1(c + \varepsilon_{t-1}) + \phi_1^2(c + \varepsilon_{t-2}) + \phi_1^3(c + \varepsilon_{t-3}) + \dots \\ &= \frac{c}{1 - \phi_1} + \varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_1^2\varepsilon_{t-2} + \phi_1^3\varepsilon_{t-3} \dots \end{aligned}$$

which follows from the limits for infinite series

Taking expectations removes everything but the first term:

$$\mathrm{E}(y_t) = \frac{c}{1 - \phi_1}$$

Implication:
if, starting at time t and going forward to ∞,
we fix $\mathrm{x}_t\boldsymbol{\beta}$,
then y_t will converge to $\mathrm{x}_t\boldsymbol{\beta}/(1 - \phi_1)$

Interpreting AR(1) parameters

$$\text{GDP}_t = \phi_1 \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t$$
$$\text{GDP}_t = 0.9 \times \text{GDP}_{t-1} + 10 + 2 \times \text{Democracy}_t + \varepsilon_t$$

If at year t, $\text{GDP}_t = 100$ and the country is a non-democracy $\text{Democracy}_t = 0$, then:

This country is in a steady state:
it will tend to have GDP of 100 every period, with small errors from ε_t (verify this)

Interpreting AR(1) parameters

$$\begin{aligned} \text{GDP}_t &= \phi_1 \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\ \text{GDP}_t &= 0.9 \times \text{GDP}_{t-1} + 10 + 2 \times \text{Democracy}_t + \varepsilon_t \end{aligned}$$

Now suppose we make the country a democracy in period $t+1$: $\text{Democracy}_{t+1} = 1$.

The model predicts that in period $t+1$, the level of GDP will rise by $\beta = 2$, to 102.

This *appears* to be a small effect, but. . .

Interpreting AR(1) parameters

$$\mathrm{GDP}_t = \phi_1 \mathrm{GDP}_{t-1} + \beta_0 + \beta_1 \mathrm{Democracy}_t + \varepsilon_t$$
$$\mathrm{GDP}_t = 0.9 \times \mathrm{GDP}_{t-1} + 10 + 2 \times \mathrm{Democracy}_t + \varepsilon_t$$

. . . the effect accumulates, so long as $\mathrm{Democracy} = 1$

$\mathrm{E}(\hat{y}_{t+2}|x_{t+2}) = 0.9 \times 102 + 10 + 2 = 103.8$

Interpreting AR(1) parameters

$$\begin{aligned} \text{GDP}_t &= \phi_1 \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\ \text{GDP}_t &= 0.9 \times \text{GDP}_{t-1} + 10 + 2 \times \text{Democracy}_t + \varepsilon_t \end{aligned}$$

. . . the effect accumulates, so long as $\text{Democracy} = 1$

$\text{E}(\hat{y}_{t+2}|x_{t+2}) = 0.9 \times 102 + 10 + 2 = 103.8$

$\text{E}(\hat{y}_{t+3}|x_{t+3}) = 0.9 \times 103.8 + 10 + 2 = 105.42$

Interpreting AR(1) parameters

$$\begin{aligned} \text{GDP}_t &= \phi_1 \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\ \text{GDP}_t &= 0.9 \times \text{GDP}_{t-1} + 10 + 2 \times \text{Democracy}_t + \varepsilon_t \end{aligned}$$

. . . the effect accumulates, so long as $\text{Democracy} = 1$

$\text{E}(\hat{y}_{t+2}|x_{t+2}) = 0.9 \times 102 + 10 + 2 = 103.8$

$\text{E}(\hat{y}_{t+3}|x_{t+3}) = 0.9 \times 103.8 + 10 + 2 = 105.42$

$\text{E}(\hat{y}_{t+4}|x_{t+4}) = 0.9 \times 105.42 + 10 + 2 = 106.878$

Interpreting AR(1) parameters

$$\begin{aligned}\text{GDP}_t &= \phi_1 \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\ \text{GDP}_t &= 0.9 \times \text{GDP}_{t-1} + 10 + 2 \times \text{Democracy}_t + \varepsilon_t\end{aligned}$$

. . . the effect accumulates, so long as $\text{Democracy} = 1$

$\text{E}(\hat{y}_{t+2}|x_{t+2}) = 0.9 \times 102 + 10 + 2 = 103.8$

$\text{E}(\hat{y}_{t+3}|x_{t+3}) = 0.9 \times 103.8 + 10 + 2 = 105.42$

$\text{E}(\hat{y}_{t+4}|x_{t+4}) = 0.9 \times 105.42 + 10 + 2 = 106.878$

. . .

$\text{E}(\hat{y}_{t=\infty}|x_{t=\infty}) = (10 + 2)/(1 - 0.9) = 120$

So is this a big effect or a small effect?

Interpreting AR(1) parameters

$\mathrm{E}(\hat{y}_{t=\infty}|x_{t=\infty}) = (10+2)/(1-0.9) = 120$

So is this a big effect or a small effect?

It depends on the length of time your covariates remain fixed.

Many comparative politics variables change rarely, so their effects accumulate slowly over time (e.g., institutions)

Presenting only β_1, rather than the accumulated change in y_t after x_t changes, could drastically *understate* the relative substantive importance of our comparative political covariates compared to rapidly changing covariates

This understatement gets larger the closer ϕ_1 gets to 1
—which is where our ϕ_1's tend to be!

Interpreting AR(1) parameters

Recommendation:
Simulate the change in y_t given a change in x_t through enough periods to capture the real-world impact of your variables

If you are studying partisan effects, and new parties tend to stay in power 5 years, don't report β_1 or the one-year change in y. Iterate out to five years.

What is the confidence interval around these cumulative changes in y given a permanent change in x?

A complex function of the se's of ϕ and $\boldsymbol{\beta}$

So simulate out to y_{t+k} using draws from the estimated distributions of $\hat{\phi}$ and $\hat{\boldsymbol{\beta}}$

`R` will help with this, using `predict()` and (in `simcf`), `ldvsimev()`

Example: UK vehicle accident deaths

Number of monthly deaths and serious injuries in UK road accidents

Data range from January 1969 to December 1984.

In February 1983, a new law requiring seat belt use took effect

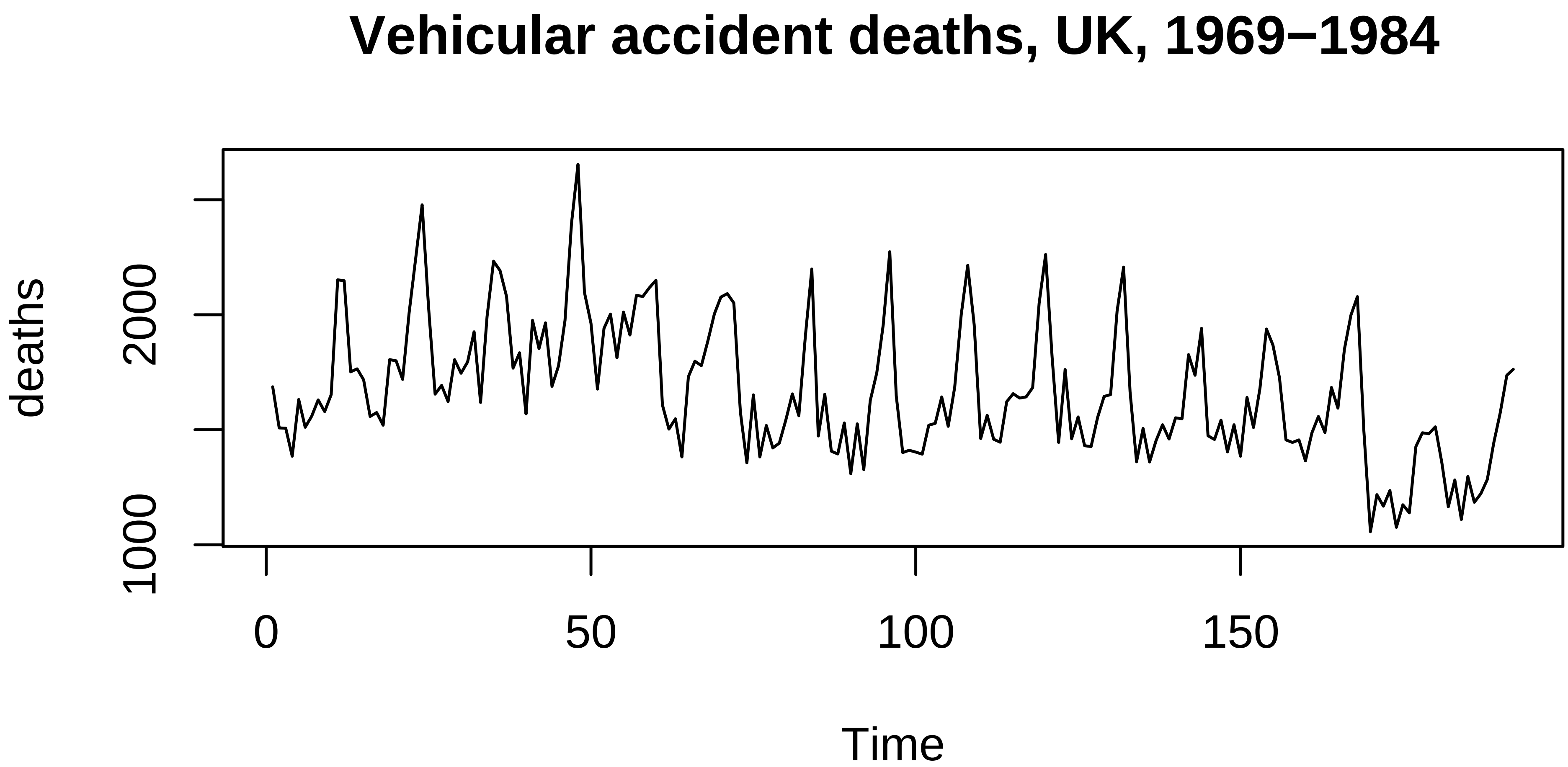
Source: Harvey, 1989, p.519ff.

`http://www.staff.city.ac.uk/s̃c397/courses/3ts/datasets.html`

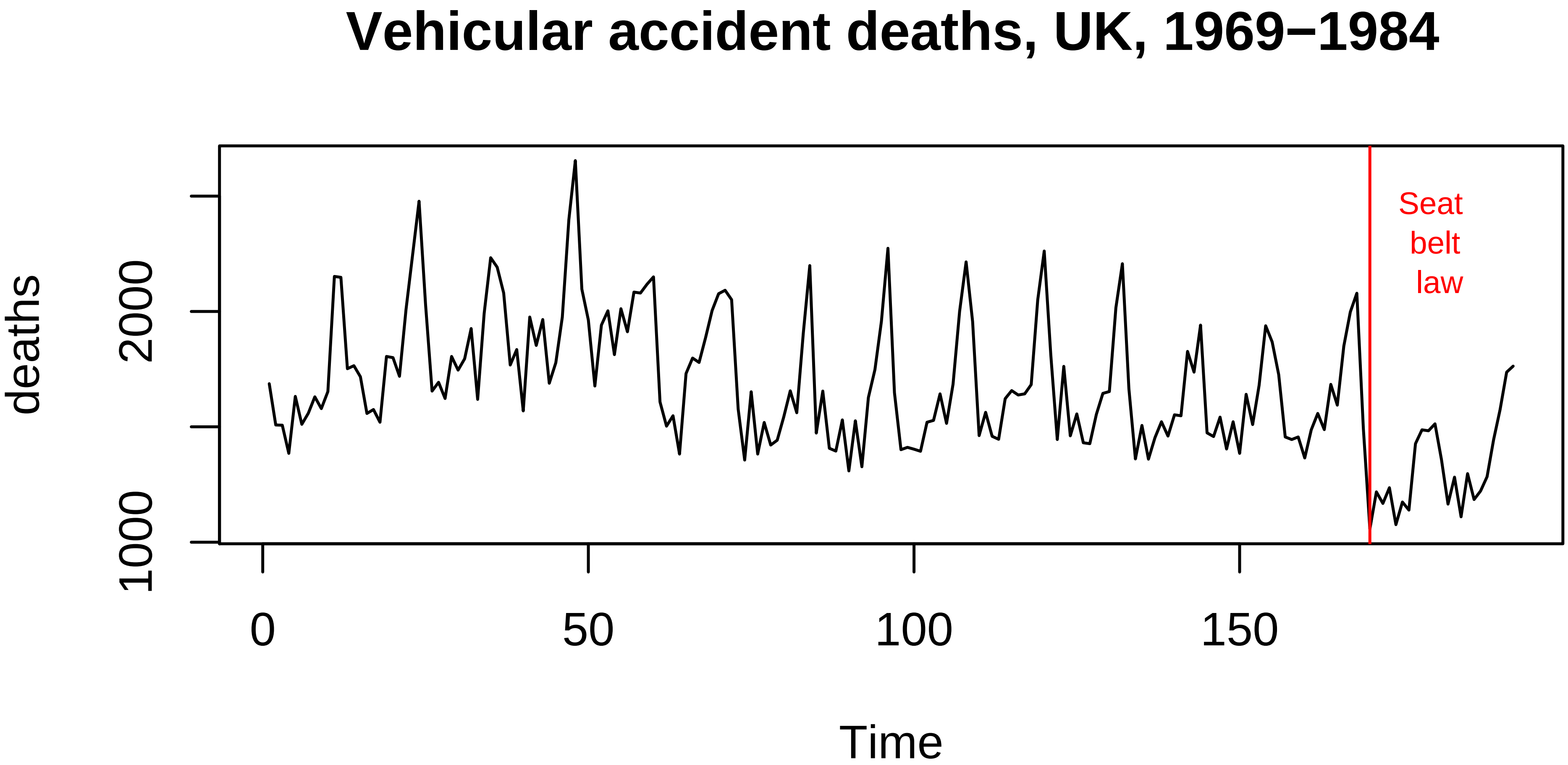
Simple, likely stationary data

Simplest possible covariate: a single dummy

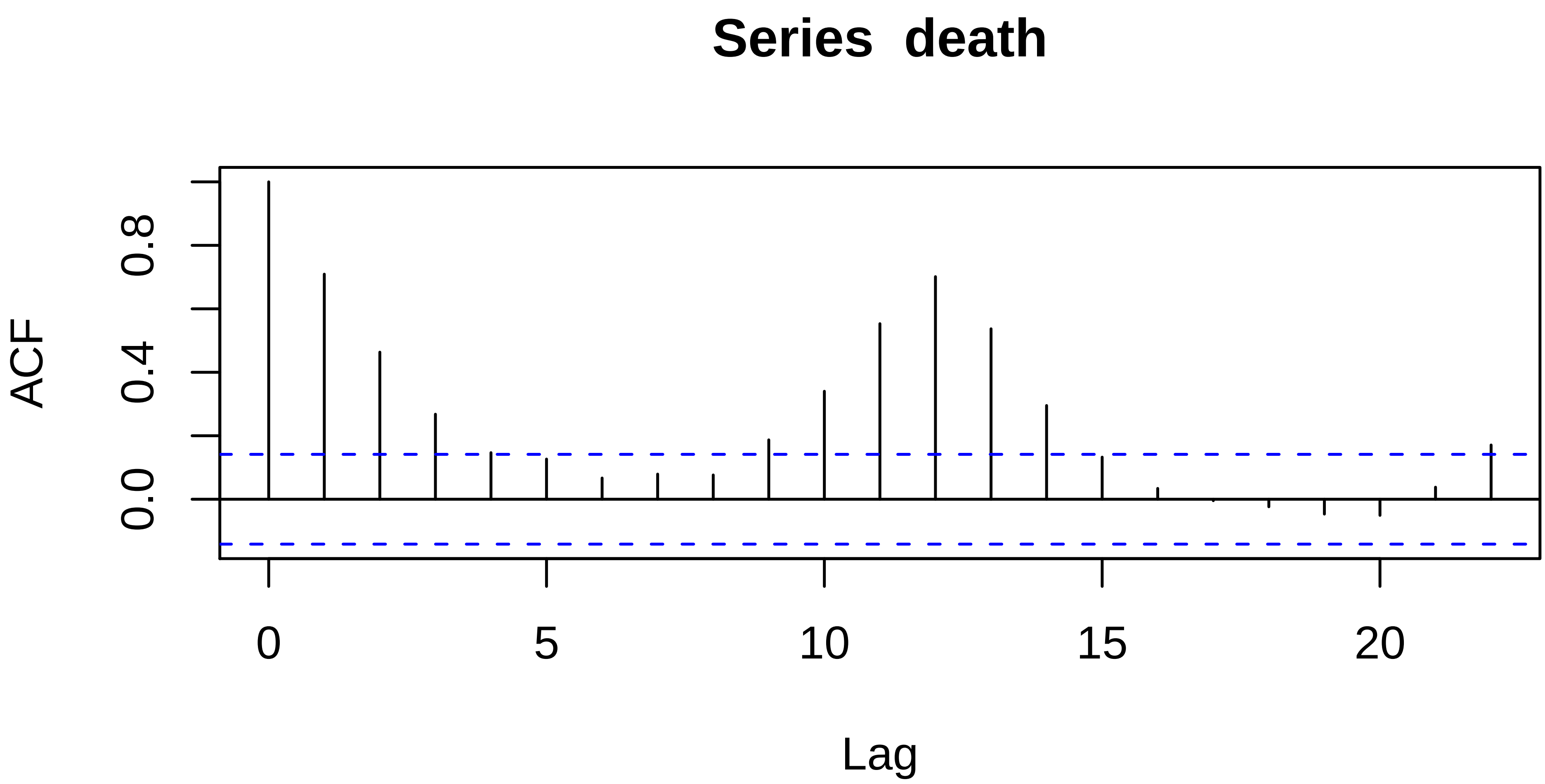
The time series



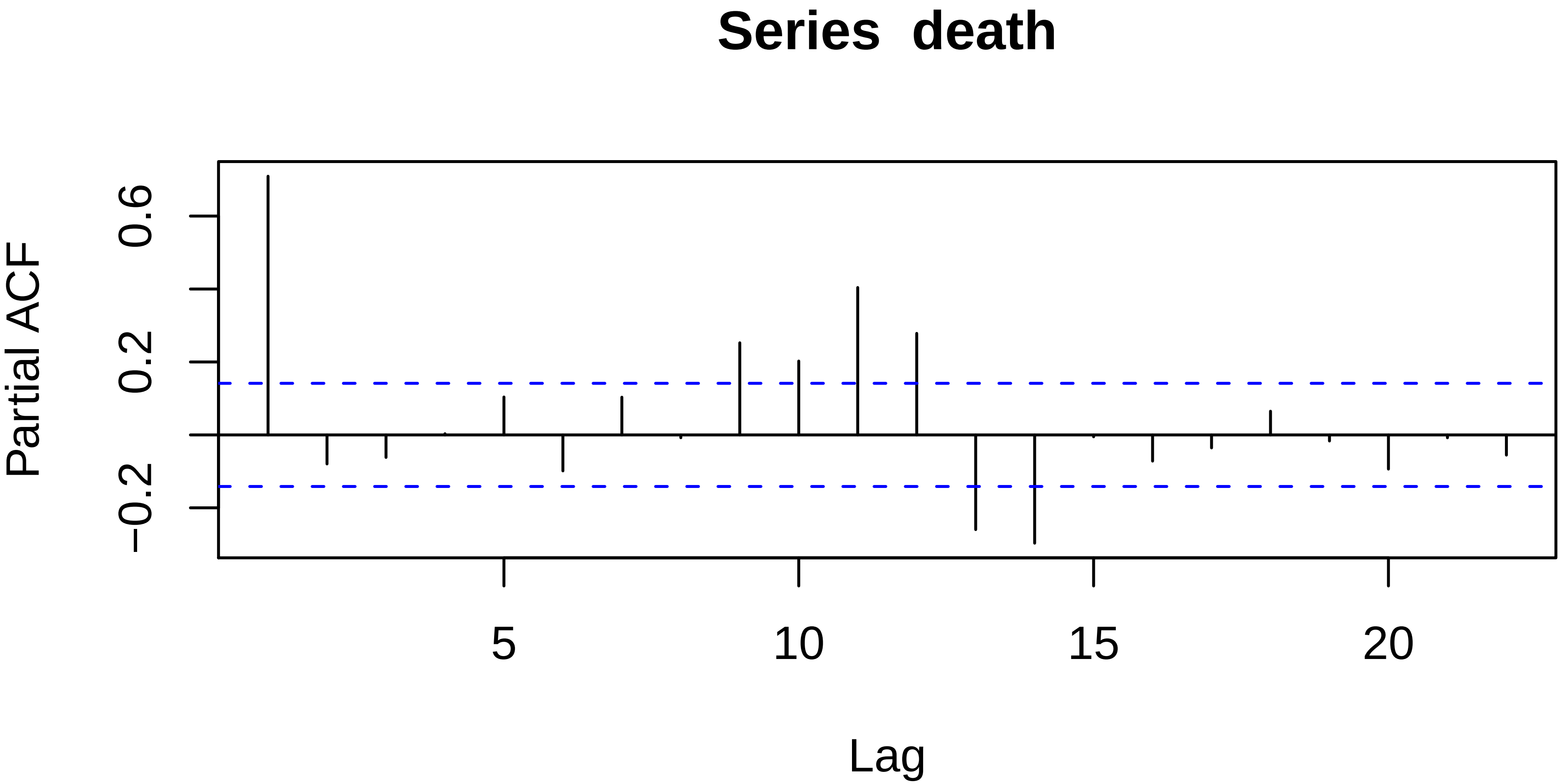
The time series



ACF



Partial ACF



AR(1) specification

```
## Estimate an AR(1) using arima
xcovariates <- law
arima.res1a <- arima(death, order = c(1,0,0),
                     xreg = xcovariates, include.mean = TRUE
                     )



Coefficients:
        ar1  intercept  xcovariates
      0.644    1719.19       -377.5
s.e.  0.055      42.08        107.7

sigma^2 estimated as 39289:  log likelihood = -1288,  aic = 2585
```

AR(1) specification with Q4 control

```
## Estimate an AR(1) using arima
xcovariates <- c(q4,law)
arima.res1a <- arima(death, order = c(1,0,0),
                     xreg = xcovariates, include.mean = TRUE
                     )
```

```
Coefficients:
        ar1  intercept     q4     law
      0.535    1638.03  324.6  -395.7
s.e.  0.064      28.12   34.5    72.3

sigma^2 estimated as 26669:  log likelihood = -1251,  aic = 2512
```

What is the effect of adding the law?

In period $t+1$? $t+12$? $t+60$

How “significant” is this effect over those periods?

An AR(p) Regression Model

The AR(p) regression model is a straightforward extension of the AR(1)

$$y_t = y_{t-1}\phi_1 + y_{t-2}\phi_2 + \ldots + y_{t-p}\phi_p + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

Note that for fixed mean, y_t now converges to

$$\mathrm{E}(y_t) = \frac{c}{1 - \phi_1 - \phi_2 - \phi_3 - \ldots - \phi_p}$$

Implication:
if, starting at time t and going forward to ∞, we fix $\mathrm{x}_i\boldsymbol{\beta}$,
then y_t will converge to $\mathrm{x}_i\boldsymbol{\beta}/(1 - \phi_1 - \phi_2 - \phi_3 - \ldots - \phi_p)$

Estimation and interpretation similar to above & uses same `R` functions

MA(1) Models

To create a regression model from the MA(1):

$$y_t = \varepsilon_{t-1}\rho_1 + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

Estimation is again by maximum likelihood

Once again a complex procedure, but still a generalization of the Normal case

Any dynamic effects in this model are quickly mean reverting

ARMA(p,q): Putting it all together

To create a regression model from the ARMA(p,q):

$$y_t = y_{t-1}\phi_1 + y_{t-2}\phi_2 + \ldots + y_{t-p}\phi_p + \varepsilon_{t-1}\rho_1 + \varepsilon_{t-2}\rho_2 + \ldots + \varepsilon_{t-q}\rho_q + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

Will need a MLE to obtain $\hat{\boldsymbol{\phi}}$, $\hat{\boldsymbol{\rho}}$, and $\hat{\boldsymbol{\beta}}$

Once again a complex procedure, but still a generalization of the Normal case

Note the AR(p) process dominates in two senses:

- Stationarity determined just by AR(p) part of ARMA(p,q)
- Long-run level determined just by AR(p) terms: still $\mathbf{x}_i\boldsymbol{\beta}/(1 - \sum_p \phi_p)$

AR(1,1) specification: Model 1c

```
xcovariates <- cbind(q4,law)
arima.res1c <- arima(death, order = c(1,0,1),
                     xreg = xcovariates, include.mean = TRUE
                     )


Coefficients:
         ar1      ma1  intercept       q4       law
       0.958  -0.768    1619.48   391.64   -384.56
s.e.   0.029   0.075      59.38    26.28     85.92

sigma^2 estimated as 24572:  log likelihood = -1243,  aic = 2499
```

AR(1,2) specification: Model 1d

```
xcovariates <- cbind(q4,law)
arima.res1d <- arima(death, order = c(1,0,2),
                     xreg = xcovariates, include.mean = TRUE
                     )



Coefficients:
         ar1      ma1      ma2  intercept       q4       law
       0.965   -0.665   -0.133     1622.1   378.19   -377.03
s.e.   0.023    0.076    0.067       61.8    28.67     85.58

sigma^2 estimated as 24097:  log likelihood = -1241,  aic = 2497
```

AR(1,3) specification: Model 1e

```
xcovariates <- cbind(q4,law)
arima.res1e <- arima(death, order = c(1,0,3),
                     xreg = xcovariates, include.mean = TRUE
                     )



Coefficients:
        ar1     ma1     ma2     ma3  intercept      q4      law
      0.967  -0.637  -0.102  -0.067     1623.7  371.57  -373.97
s.e.  0.022   0.083   0.078   0.073       61.9   30.16    86.58

sigma^2 estimated as 23995:  log likelihood = -1241,  aic = 2498
```

AR(1,3) specification: Model 1f

```
xcovariates <- cbind(q4,law)
arima.res1f <- arima(death, order = c(2,0,1),
                     xreg = xcovariates, include.mean = TRUE
                     )


Coefficients:
         ar1     ar2     ma1  intercept     q4      law
       1.155  -0.182  -0.840    1622.53  374.31  -375.38
s.e.   0.098   0.091   0.054      61.92   30.00    86.11

sigma^2 estimated as 24060:  log likelihood = -1241,  aic = 2497
```

Selected model 1: ARMA(1,2)

```
xcovariates <- cbind(q4,law)
arima.res1d <- arima(death, order = c(1,0,2),
                     xreg = xcovariates, include.mean = TRUE
                     )
```

```
Coefficients:
        ar1     ma1     ma2  intercept      q4      law
      0.965  -0.665  -0.133     1622.1  378.19  -377.03
s.e.  0.023   0.076   0.067       61.8   28.67    85.58

sigma^2 estimated as 24097:  log likelihood = -1241,  aic = 2497
```

What does this mean?

Where does this series go in the limit?

Counterfactual forecasting

1. Start in period t with the observed y_t and $\mathbf{x}_t$

Counterfactual forecasting

1. Start in period t with the observed y_t and $\mathbf{x}_t$

2. Choose hypothetical $\mathbf{x}_{c,t}$ for every period t to $t+k$ you wish to forecast

Counterfactual forecasting

1. Start in period t with the observed y_t and $\mathbf{x}_t$

2. Choose hypothetical $\mathbf{x}_{c,t}$ for every period t to $t+k$ you wish to forecast

3. Estimate $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, $\boldsymbol{\rho}$, and σ^2

Counterfactual forecasting

1. Start in period t with the observed y_t and $\mathbf{x}_t$
2. Choose hypothetical $\mathbf{x}_{c,t}$ for every period t to $t + k$ you wish to forecast
3. Estimate $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, $\boldsymbol{\rho}$, and σ^2
4. Draw a vector of these parameters from their predictive distribution as estimated by the MLE

Counterfactual forecasting

1. Start in period t with the observed y_t and $\mathbf{x}_t$
2. Choose hypothetical $\mathbf{x}_{c,t}$ for every period t to $t+k$ you wish to forecast
3. Estimate $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, $\boldsymbol{\rho}$, and σ^2
4. Draw a vector of these parameters from their predictive distribution as estimated by the MLE
5. Calculate one simulated value of $\tilde{y}$ for the next step, $t+1$, using:

$$\tilde{y}_{t+1} = \sum_p y_{t-p}\tilde{\phi}_p + \mathbf{x}_{c,t+1}\tilde{\boldsymbol{\beta}} + \sum_q \varepsilon_{t+q}\tilde{\rho}_q + \tilde{\varepsilon}_t$$

Counterfactual forecasting

1. Start in period t with the observed y_t and $\mathbf{x}_t$
2. Choose hypothetical $\mathbf{x}_{c,t}$ for every period t to $t+k$ you wish to forecast
3. Estimate $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, $\boldsymbol{\rho}$, and σ^2
4. Draw a vector of these parameters from their predictive distribution as estimated by the MLE
5. Calculate one simulated value of $\tilde{y}$ for the next step, $t+1$, using:

$$\tilde{y}_{t+1} = \sum_{p} y_{t-p}\tilde{\phi}_p + \mathbf{x}_{c,t+1}\tilde{\boldsymbol{\beta}} + \sum_{q} \varepsilon_{t+q}\tilde{\rho}_q + \tilde{\varepsilon}_t$$

6. Move to the next period, $t+2$, and using the past actual *and* forecast values of y_t and ε_t as lags; repeat until you reach period $t+k$

7. You have one simulated forecast. Repeat steps 4–6 until you have many (say, 1000) simulated forecasts. These are your predicted values, and can be summarized by a mean and predictive intervals

7. You have one simulated forecast. Repeat steps 4–6 until you have many (say, 1000) simulated forecasts. These are your predicted values, and can be summarized by a mean and predictive intervals

8. To get a simulated expected forecast, repeat step 7 many times (say 1000), each time taking the average forecast. You now have a vector of 1000 expected forecasts, and can summarize them with a mean and confidence intervals

Effect of repealing seatbelt law?

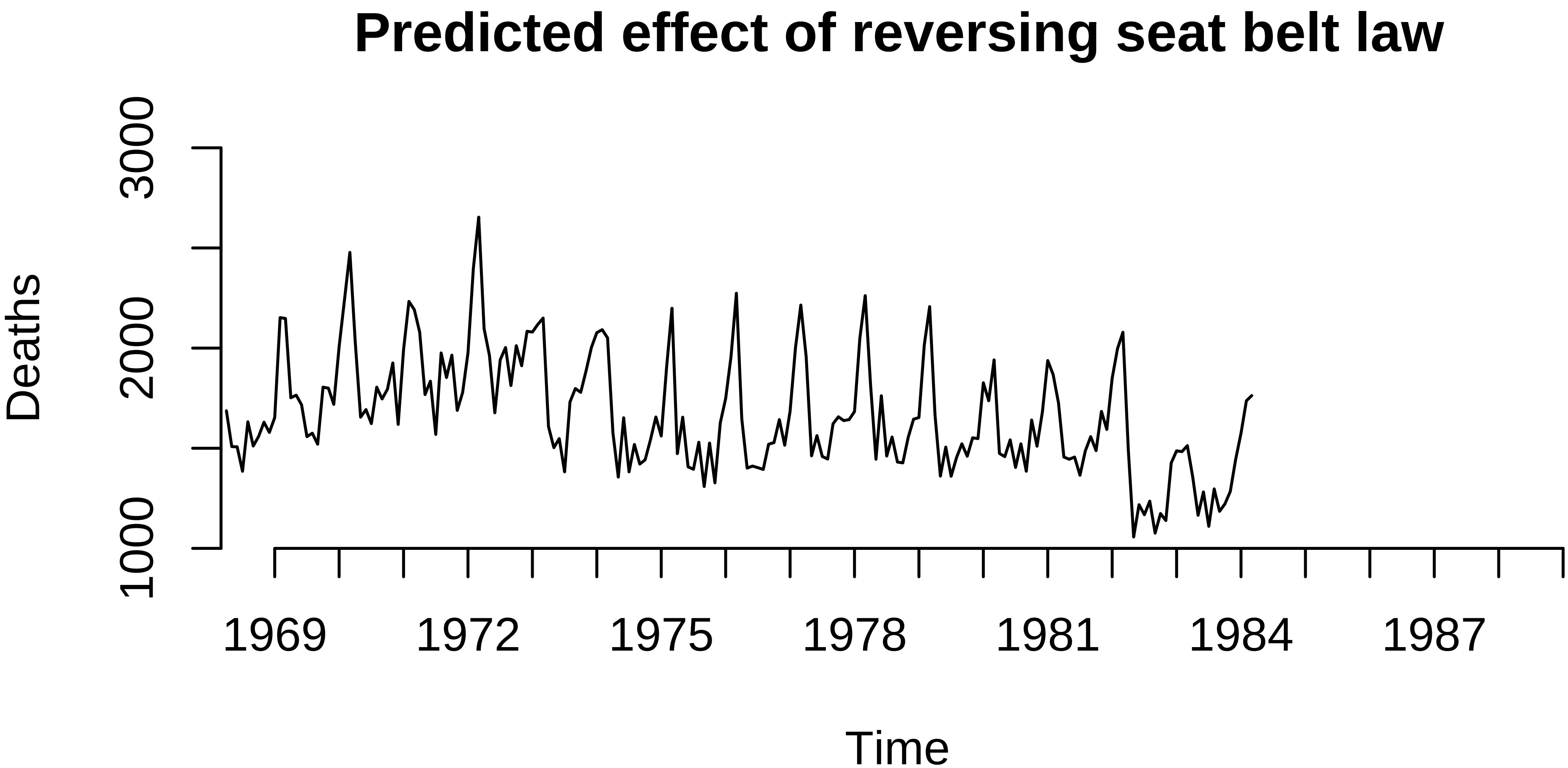
What does the model predict would happen if we repealed the law?

How much would deaths increase after one month? One year? Five years?

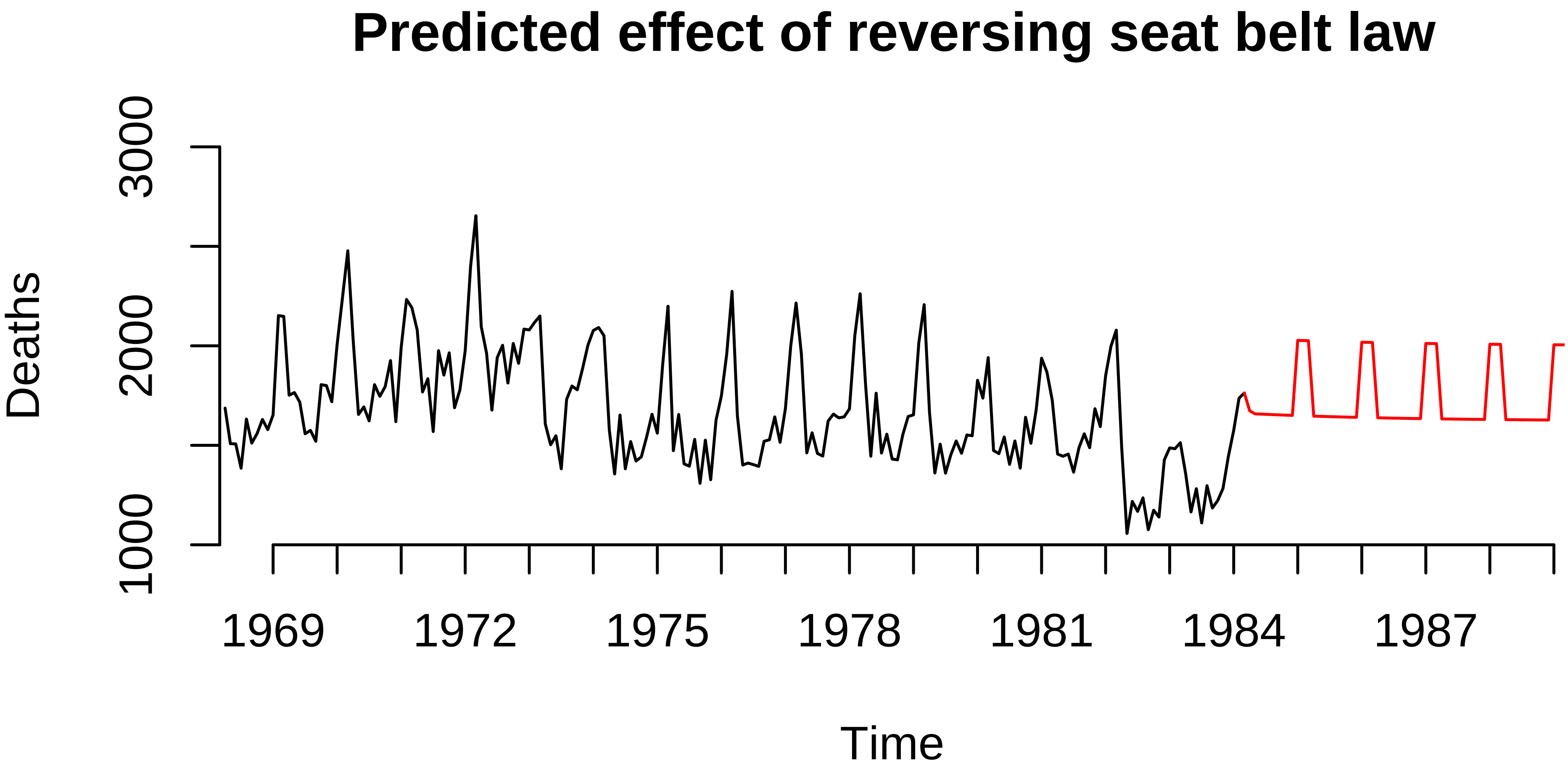
If we run this experiment, how much might the results vary from model expectations?

Need forecast deaths—no law for the next 60 periods, plus predictive intervals

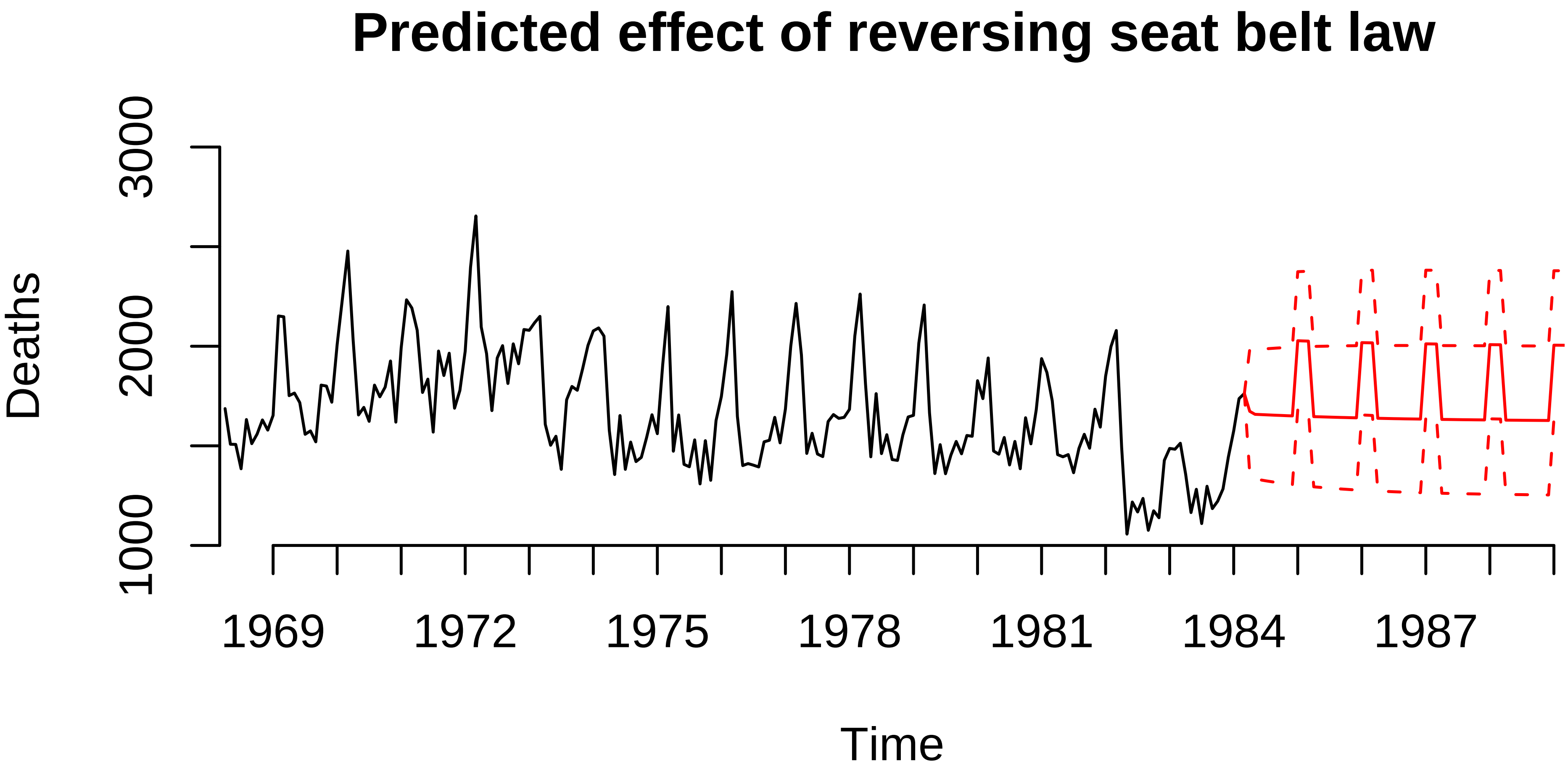
```
predict(arima.res1,                # The model
        n.ahead = 60,              # predict out 60 periods
        newxreg = newdata)         # using these counterfactual x's
```



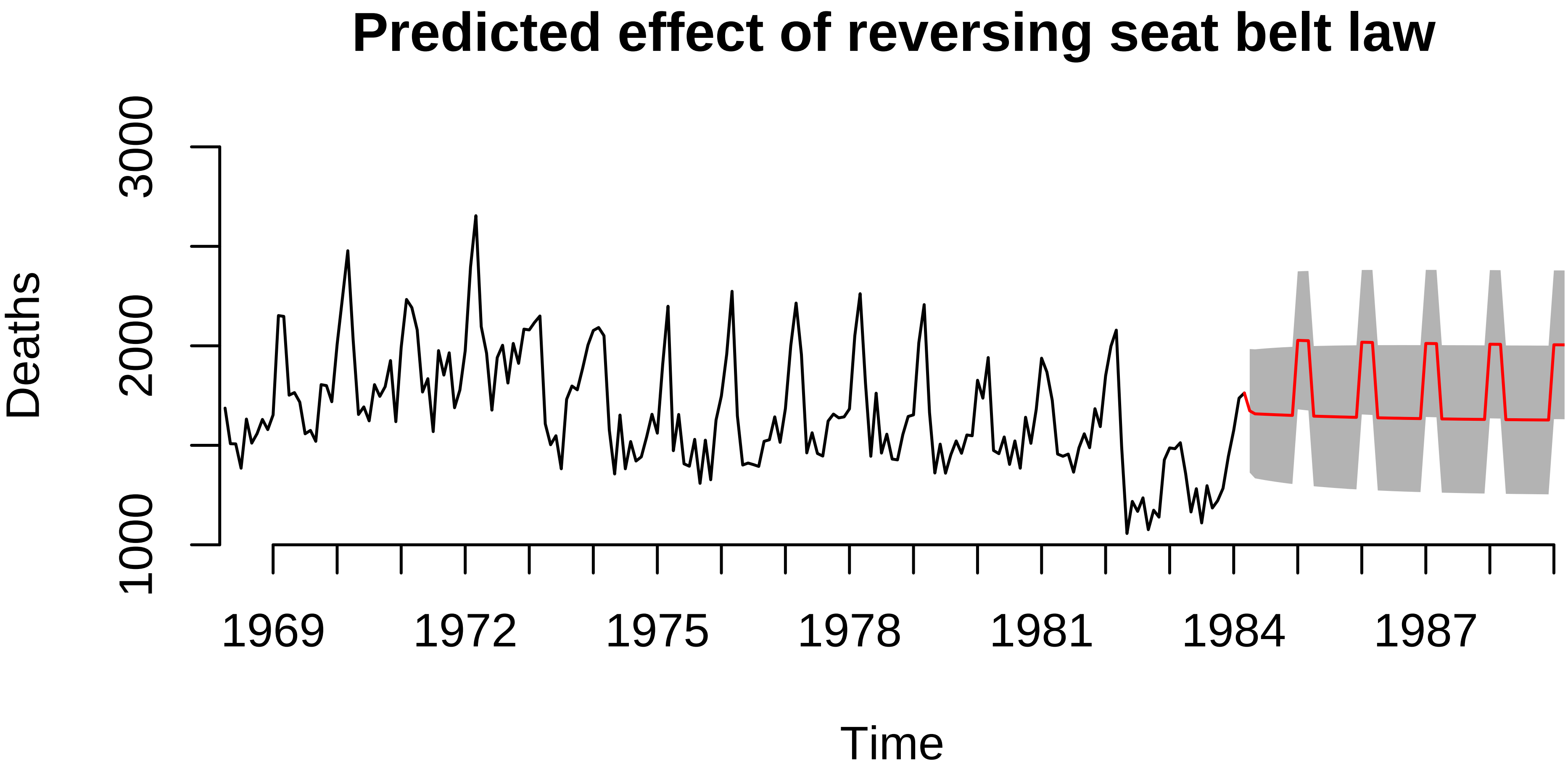
The observed time series



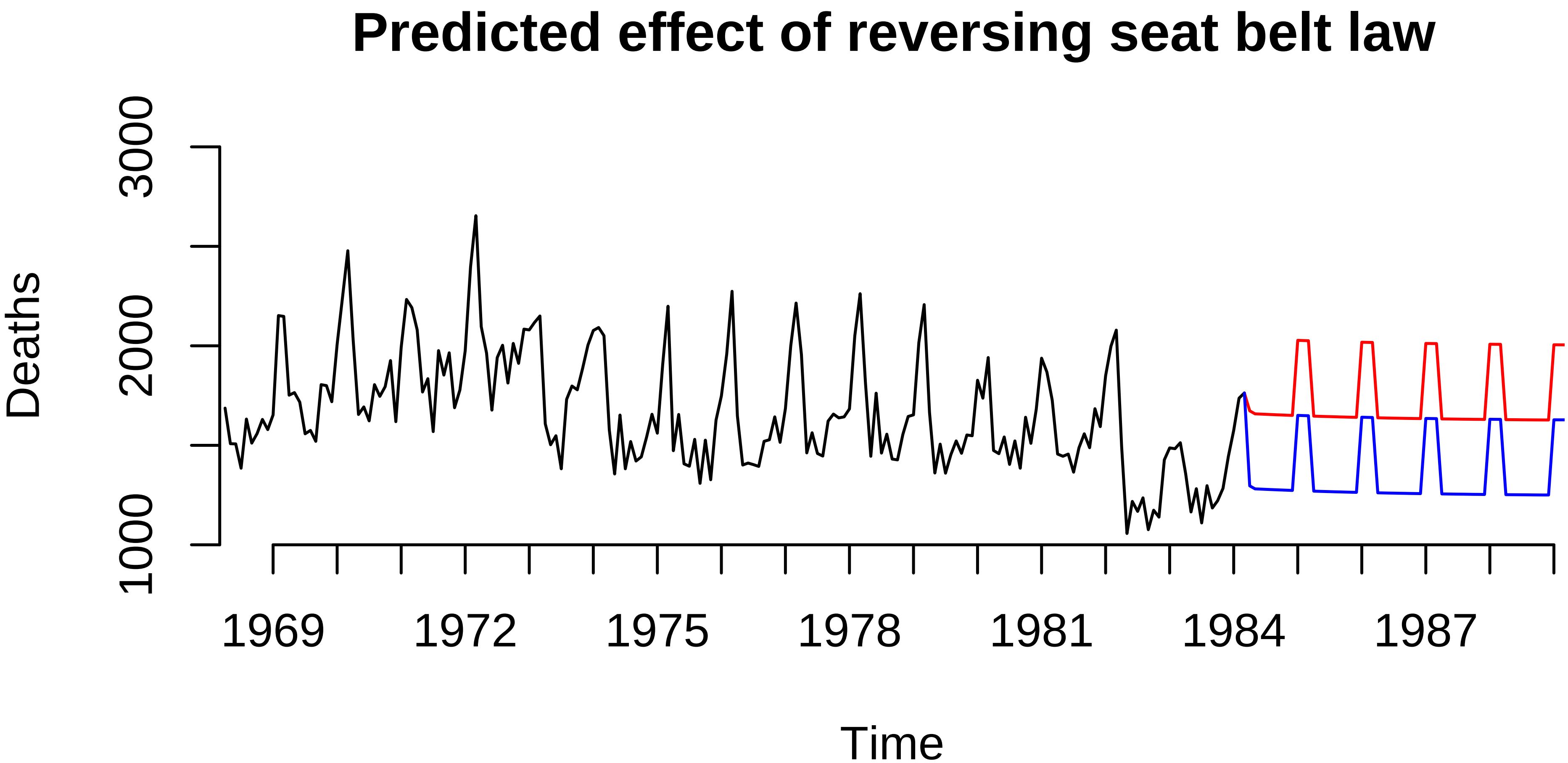
What the model predicts would happen if the seat belt requirement is *repealed*



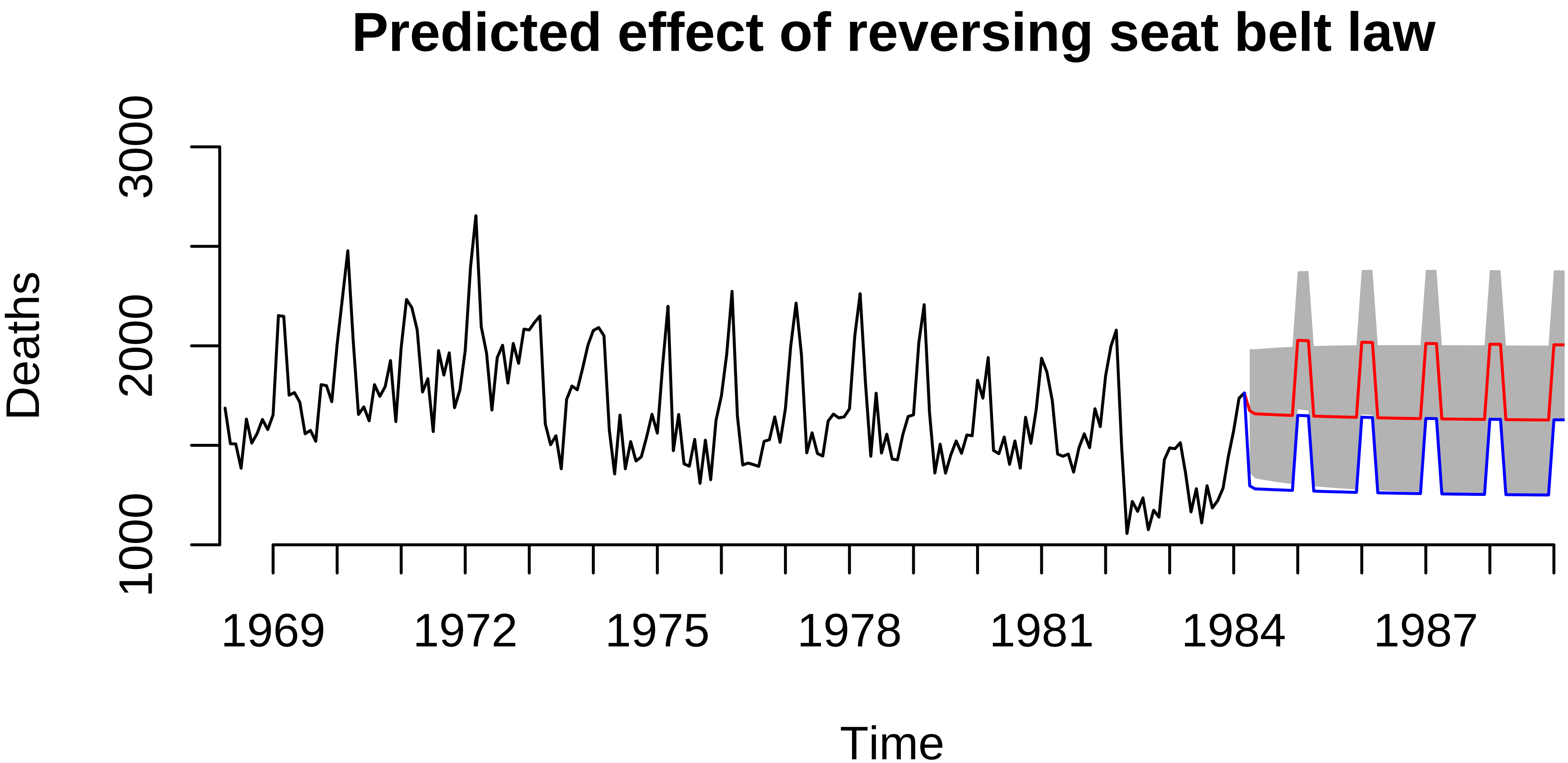
adding the 95 % predictive interval



which is easier to read as a polygon



comparing to what would happen with the law left intact



comparing to what would happen with the law left intact

Confidence intervals vs. Predictive Intervals

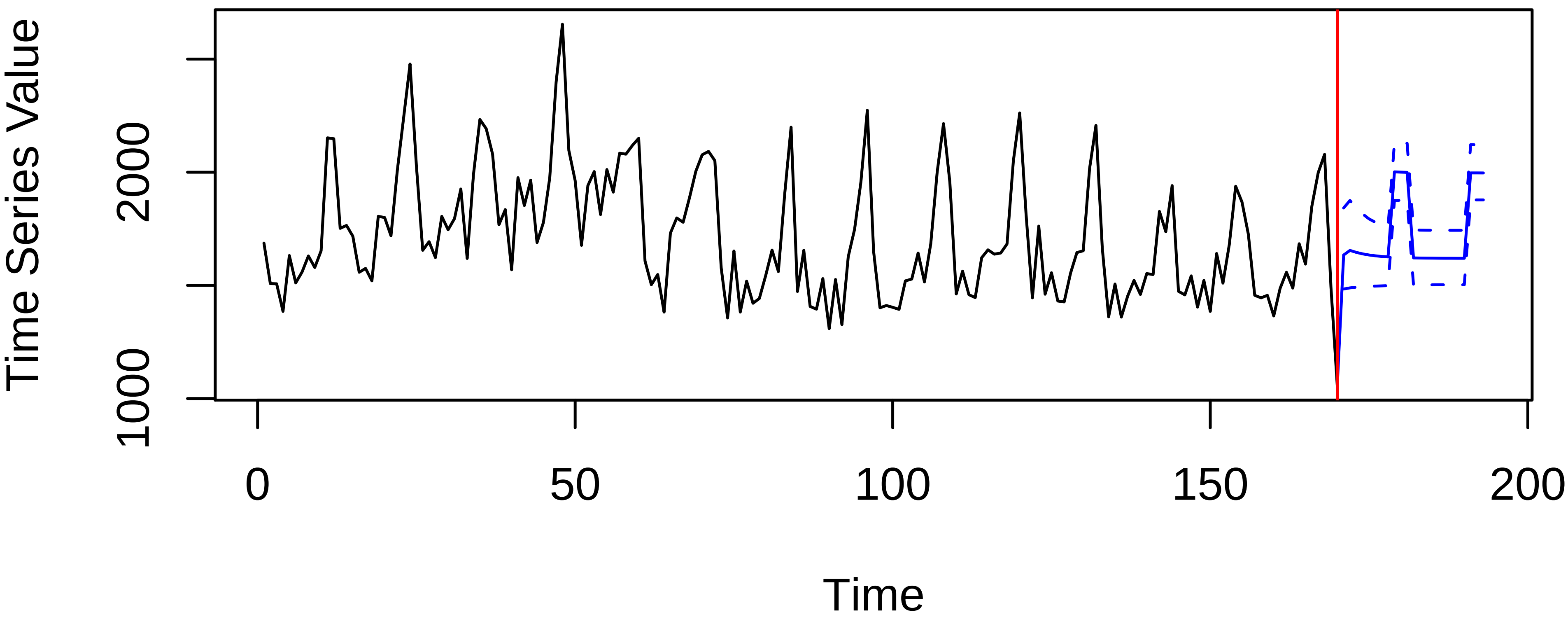
Suppose we want *confidence intervals* instead of *predictive intervals*

CIs just show the uncertainty from estimation

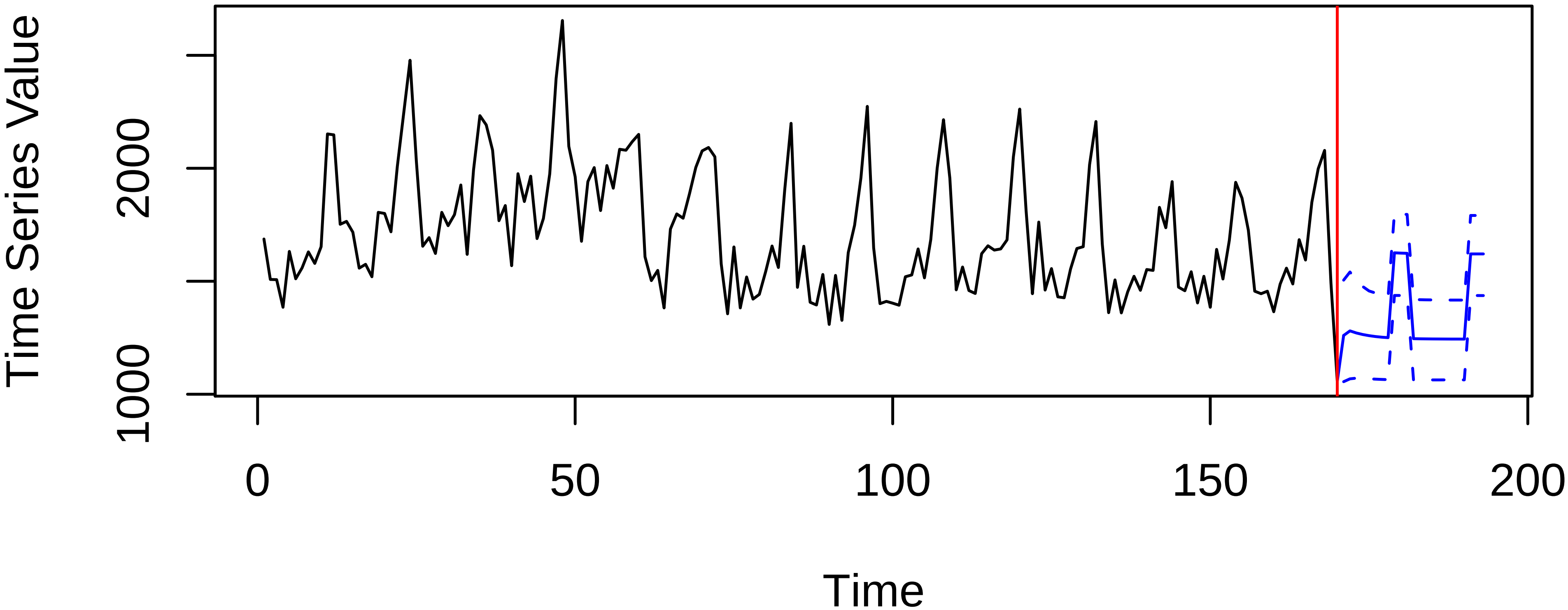
Analog to $\mathrm{se}(\beta)$ and significance tests

`predict()` *won't* give us CIs

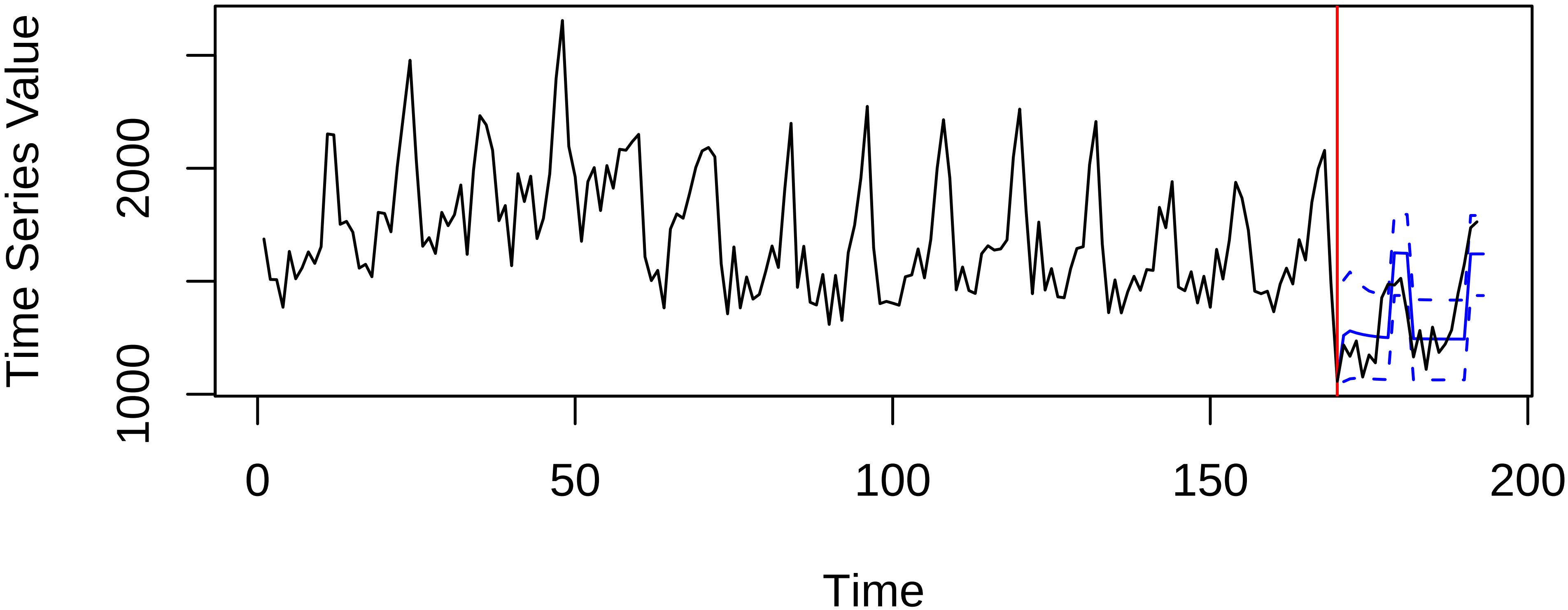
Need to use another package, `Zelig`. (Will review code later.)



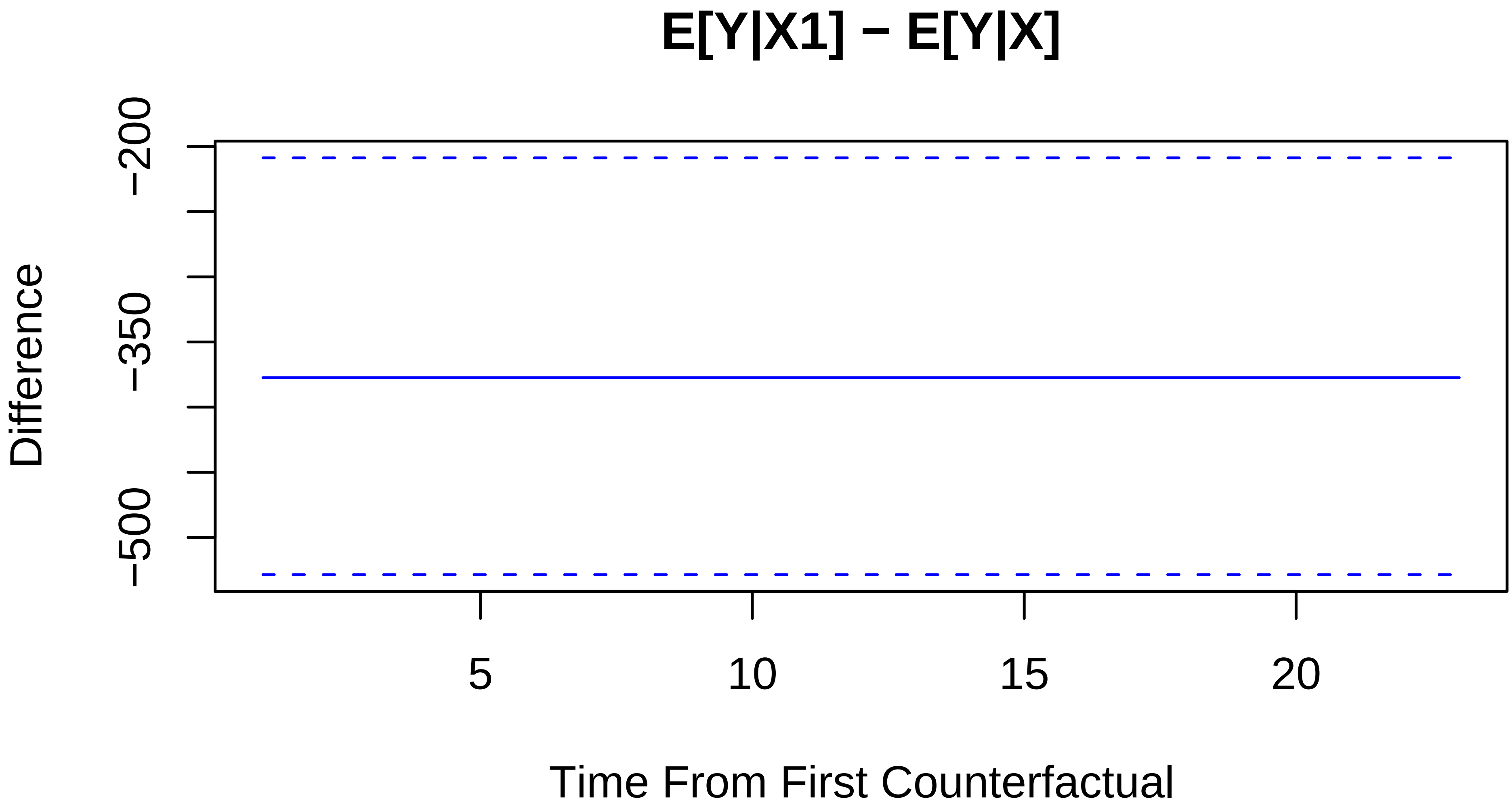
The blue lines show what the model predicts would have happened if no seat belt law had been implemented. Dashed lines are 95% *confidence* intervals around the expected number of deaths given the lack of a law



The blue lines now show what the model predicts should have happened under the (factual) scenario in which a law was implemented.



The model expectations fit closely with the actual data



The model estimates a large, statistically significant and constant reduction in deaths due to the law. Won't always be constant

Neat. But is ARMA(p,q) appropriate for our data?

ARMA(p,q) an extremely flexible, broadly applicable model of single time series y_t

But ONLY IF y_t is stationary

If data are non-stationary (have a unit root), then:

- Results may be spurrious
- Long-run predictions impossible

Can assess stationarity through two methods:

1. Examine the data: time series, ACF, and PACF plots
2. Statistical tests for a unit root

Unit root tests: Basic notion

- If y_t is stationary, large negative shifts should be followed by large positive shifts, and vice versa (mean-reversion)

- If y_t is non-stationary (has a unit root), large negative shifts should be uncorrelated with large positive shifts

Thus if we regress $y_t - y_{t-1}$ on y_{t-1}, we should get a negative coefficient if and only if the series is stationary

To do this:

Augmented Dickey-Fuller test `adf.test()` in the `tseries` library

Phillips-Perron test: `PP.test()`

Tests differ in how they model heteroskedasticity, serial correlation, and the number of lags

Unit root tests: Limitations

Form of unit root test: rejecting the null of a unit root.

Will tend to fail to reject for many non-unit roots with high persistence

Very hard to distinguish near-unit roots from unit roots with test statistics

Famously low power tests

Unit root tests: Limitations

Analogy: Using polling data to predict a very close election

Null Hypothesis: Left Party will get 50.01% of the vote

Alternative Hypothesis: Left will get $< 50\%$ of the vote

We're okay with a 3% CI if we're interested in alternatives like 45% of the vote

But suppose we need to compare the Null to 49.99%

To confidently reject the Null in favor of a very close alternative like this, we'd need a CI of about 0.005% or less

Unit root tests: Limitations

In comparative politics, we usual ask whether $\phi = 1$ or, say, $\phi = 0.99$

Small numerical difference makes a huge difference for modeling

And unit root tests are weak, and poorly discriminate across these cases

Simply not much use to us

Unit root tests: usage

```
> # Check for a unit root
> PP.test(death)

        Phillips-Perron Unit Root Test

data:  death
Dickey-Fuller = -6.435, Truncation lag parameter = 4, p-value = 0.01

> adf.test(death)

        Augmented Dickey-Fuller Test

data:  death
Dickey-Fuller = -6.537, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

Linear regression with $\mathbf{Y}_{t-1}$

A popular model in comparative politics is:

$$y_t = y_{t-1}\phi_1 + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

estimated by least squares, rather than maximum likelihood

That is, treat y_{t-1} as "just another covariate", rather than a special term

Danger of this approach: y_{t-1} and ε_t are almost certainly correlated

Violates G-M condition 3: Bias in $\boldsymbol{\beta}$, incorrect s.e.'s

When can you use a lagged y?

My recommendation:

1. Estimate an LS model with the lagged DV
2. Check for remaining serial correlation (Breusch-Godfrey)
3. Compare your results to the corresponding AR(p) estimated by ML
4. Use LS only if it make no statistical or substantive difference

Upshot: You can use LS in cases where it works just as well as ML

If you model the right number of lags, and need no MA(q) terms, LS often not far off

Still need to interpret the β's and ϕ's dynamically

Testing for serial correlation in errors

In LS models, serial correlation makes estimates inefficient (like heteroskedasticity)

Testing for serial correlation in errors

In LS models, serial correlation makes estimates inefficient (like heteroskedasticity)

If the model includes a lagged dependent variable, serial correlation $\rightarrow$ inconsistent estimates $(\mathrm{E}(x\epsilon) \neq 0)$

Testing for serial correlation in errors

In LS models, serial correlation makes estimates inefficient (like heteroskedasticity)

If the model includes a lagged dependent variable, serial correlation $\rightarrow$ inconsistent estimates $(\mathrm{E}(x\epsilon) \neq 0)$

So we need to be able to test for serial correlation.

A general test that will work for single time series or panel data is based on the Lagrange Multiplier

Called Breusch-Godfrey test, or the LM test

Lagrange Multiplier test for serial correlation

1. Run your time series regression by least squares, regressing

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt} + \phi_1 y_{t-1} + \ldots + \phi_k y_{t-k} + u_t$$

Lagrange Multiplier test for serial correlation

1. Run your time series regression by least squares, regressing

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt} + \phi_1 y_{t-1} + \ldots + \phi_k y_{t-k} + u_t$$

2. Regress (using LS) $\hat{u}_t$ on a constant,
   the explanatory variables $x_1, \ldots, x_k, y_{t-1}, \ldots, y_{t-k}$,
   and the lagged residuals, $\hat{u}_{t-1}, \ldots \hat{u}_{t-m}$

   Be sure to chose $m < p$. If you choose $m = 1$, you have a test for 1st degree autocorrelation; if you choose $m = 2$, you have a test for 2nd degree autocorrelation, etc.

Lagrange Multiplier test for serial correlation

1. Run your time series regression by least squares, regressing

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt} + \phi_1 y_{t-1} + \ldots + \phi_k y_{t-k} + u_t$$

2. Regress (using LS) $\hat{u}_t$ on a constant,
   the explanatory variables $x_1, \ldots, x_k, y_{t-1}, \ldots, y_{t-k}$,
   and the lagged residuals, $\hat{u}_{t-1}, \ldots \hat{u}_{t-m}$

   Be sure to chose $m < p$. If you choose $m = 1$, you have a test for 1st degree autocorrelation; if you choose $m = 2$, you have a test for 2nd degree autocorrelation, etc.

3. Compute the test-statistic $(T - p)R^2$, where R^2 is the coefficient of determination from the regression in step 2. This test statistic is distributed χ^2 with m degrees of freedom.

Lagrange Multiplier test for serial correlation

1. Run your time series regression by least squares, regressing

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt} + \phi_1 y_{t-1} + \ldots + \phi_k y_{t-k} + u_t$$

2. Regress (using LS) $\hat{u}_t$ on a constant,
   the explanatory variables $x_1, \ldots, x_k, y_{t-1}, \ldots, y_{t-k}$,
   and the lagged residuals, $\hat{u}_{t-1}, \ldots \hat{u}_{t-m}$

   Be sure to chose $m < p$. If you choose $m = 1$, you have a test for 1st degree autocorrelation; if you choose $m = 2$, you have a test for 2nd degree autocorrelation, etc.

3. Compute the test-statistic $(T - p)R^2$, where R^2 is the coefficient of determination from the regression in step 2. This test statistic is distributed χ^2 with m degrees of freedom.

4. Rejecting the null for this test statistic is equivalent to rejecting no autocorrelation.

Regression with lagged DV for Accidents

```
Call:
lm(formula = death ~ lagdeath + q4 + law)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  848.4006    79.4700   10.68  < 2e-16 ***
lagdeath       0.4605     0.0469    9.82  < 2e-16 ***
q4           311.5325    27.8085   11.20  < 2e-16 ***
law         -211.2391    39.8187   -5.31  3.2e-07 ***

Multiple R-squared: 0.714,      Adjusted R-squared: 0.709
```

Tests for serial correlation

```
> bgtest(lm.res1)

        Breusch-Godfrey test for serial correlation of order 1

data:  lm.res1
LM test = 0.016, df = 1, p-value = 0.8995

> bgtest(lm.res1,2)

        Breusch-Godfrey test for serial correlation of order 2

data:  lm.res1
LM test = 10.92, df = 2, p-value = 0.004259
```

What we're doing today

Next steps:

- Review ARMA(p,q) prediction and confidence intervals
- Discuss distributed lag models
- Learn some (weak) techniques for identifying non-stationary time series
- Analyze non-stationary series using differences
- Analyze non-stationary series using cointegration

Differences & Integrated time series

Define $\Delta^d y_t$ as the dth difference of y_t

For the first difference $(d = 1)$, we write

$$\Delta y_t = y_t - y_{t-1}$$

For the second difference $(d = 2)$, we write

$$\Delta^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

or the difference of two first differences

or the difference in the difference

Differences & Integrated time series

For the third difference ($d = 3$), we write

$$\Delta^3 y_t = ((y_t - y_{t-1}) - (y_{t-1} - y_{t-2})) - (y_{t-1} - y_{t-2}) - (y_{t-2} - y_{t-3})$$

or the difference of two second differences

or the difference in the difference in the difference

This gets perplexing fast.

Fortunately, we will rarely need $d > 1$, and almost never $d > 2$.

Differences & Integrated time series

What happens if we difference a stationary AR(1) process ($|\phi_1| < 1$)?

$$y_t = y_{t-1}\phi_1 + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

Differences & Integrated time series

What happens if we difference a stationary AR(1) process ($|\phi_1| < 1$)?

$$\begin{aligned} y_t &= y_{t-1}\phi_1 + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\ y_t - y_{t-1} &= y_{t-1}\phi_1 - y_{t-1} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \end{aligned}$$

Differences & Integrated time series

What happens if we difference a stationary AR(1) process ($|\phi_1| < 1$)?

$$\begin{aligned} y_t &= y_{t-1}\phi_1 + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\ y_t - y_{t-1} &= y_{t-1}\phi_1 - y_{t-1} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\ \Delta y_t &= (1-\phi)y_{t-1} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \end{aligned}$$

We still have an AR(1) process, *and* we've thrown away some useful information (the levels in y_t) that our covariates x_t might explain

Differences & Integrated time series

What happens if we difference a random walk?

$$y_t \quad = \quad y_{t-1} + \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t$$

Differences & Integrated time series

What happens if we difference a random walk?

$$y_t = y_{t-1} + \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t$$

$$y_t - y_{t-1} = y_{t-1} - y_{t-1} + \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t$$

Differences & Integrated time series

What happens if we difference a random walk?

$$\begin{aligned} y_t &= y_{t-1} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\ y_t - y_{t-1} &= y_{t-1} - y_{t-1} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\ \Delta y_t &= \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \end{aligned}$$

The result is AR(0), and stationary—
we could analyze it using ARMA(0,0), which is just LS regression!

When a single differencing removes non-stationarity from a time series y_t,
we say y_t is *integrated* of order 1, or I(1).

A time series that does not need to be differenced to be stationary is I(0).

This differencing trick comes at a price:
we can only explain changes in y_t, *not* levels,
and hence not the long-run relationship between y_t and $\mathbf{x}_t$.

Differences & Integrated time series

What happens if we difference an AR(2) unit root process?

$$y_t = 1.5y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t$$

Differences & Integrated time series

What happens if we difference an AR(2) unit root process?

$$
\begin{aligned}
y_t &= 1.5y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\
y_t - y_{t-1} &= 1.5y_{t-1} - y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t
\end{aligned}
$$

Differences & Integrated time series

What happens if we difference an AR(2) unit root process?

$$\begin{aligned} y_t &= 1.5y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\ y_t - y_{t-1} &= 1.5y_{t-1} - y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \\ \Delta y_t &= 0.5y_{t-1} - 0.5y_{t-2} + \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \end{aligned}$$

We get a stationary AR(2) process. We could analyze this new process with ARMA(2,0).

We say that the original process is ARI(2,1),
or an integrated autoregressive process of order 2, integrated of order 1.

Differences & Integrated time series

Recall our GDP & Democracy example

$$\begin{aligned} \text{GDP}_t &= \phi_1 \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\ \text{GDP}_t &= 0.9 \times \text{GDP}_{t-1} + 10 + 2 \times \text{Democracy}_t + \varepsilon_t \end{aligned}$$

At year t, $\text{GDP}_t = 100$ and the country is a non-democracy $\text{Democracy}_t = 0$, and we are curious what would happen to GDP if in $t+1$ to $t+k$, the country becomes a democracy.

Differences & Integrated time series

At year t, $\mathrm{GDP}_t = 100$ and the country is a non-democracy $\mathrm{Democracy}_t = 0$, and we are curious what would happen to GDP if in $t+1$ to $t+k$, the country becomes a democracy.

$$\mathrm{GDP}_t \quad = \quad \phi_1 \mathrm{GDP}_{t-1} + \beta_0 + \beta_1 \mathrm{Democracy}_t + \varepsilon_t$$

Differences & Integrated time series

At year t, $\text{GDP}_t = 100$ and the country is a non-democracy $\text{Democracy}_t = 0$, and we are curious what would happen to GDP if in $t+1$ to $t+k$, the country becomes a democracy.

$$\begin{aligned}
\text{GDP}_t &= \phi_1 \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\
\text{GDP}_t - \text{GDP}_{t-1} &= \phi_1 \text{GDP}_{t-1} - \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t
\end{aligned}$$

Differences & Integrated time series

At year t, $\mathrm{GDP}_t = 100$ and the country is a non-democracy $\mathrm{Democracy}_t = 0$, and we are curious what would happen to GDP if in $t+1$ to $t+k$, the country becomes a democracy.

$$
\begin{aligned}
\mathrm{GDP}_t &= \phi_1 \mathrm{GDP}_{t-1} + \beta_0 + \beta_1 \mathrm{Democracy}_t + \varepsilon_t \\
\mathrm{GDP}_t - \mathrm{GDP}_{t-1} &= \phi_1 \mathrm{GDP}_{t-1} - \mathrm{GDP}_{t-1} + \beta_0 + \beta_1 \mathrm{Democracy}_t + \varepsilon_t \\
\Delta \mathrm{GDP}_t &= (1 - \phi_1) \mathrm{GDP}_{t-1} - \mathrm{GDP}_{t-1} + \beta_0 + \beta_1 \mathrm{Democracy}_t + \varepsilon_t
\end{aligned}
$$

Differences & Integrated time series

At year t, $\text{GDP}_t = 100$ and the country is a non-democracy $\text{Democracy}_t = 0$, and we are curious what would happen to GDP if in $t+1$ to $t+k$, the country becomes a democracy.

$$\begin{aligned}
\text{GDP}_t &= \phi_1 \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\
\text{GDP}_t - \text{GDP}_{t-1} &= \phi_1 \text{GDP}_{t-1} - \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\
\Delta\text{GDP}_t &= (1 - \phi_1)\text{GDP}_{t-1} - \text{GDP}_{t-1} + \beta_0 + \beta_1 \text{Democracy}_t + \varepsilon_t \\
\Delta\text{GDP}_t &= -0.1 \times \text{GDP}_{t-1} + 10 + 2 \times \text{Democracy}_t + \varepsilon_t
\end{aligned}$$

Works just as before—but we have to supply external information on the *levels*

The model doesn't know them

ARIMA(p,d,q) models

An ARIMA(p,d,q) regression model has the following form:

$$\begin{aligned}\Delta^d y_t &= \Delta^d y_{t-1}\phi_1 + \Delta^d y_{t-2}\phi_2 + \ldots + \Delta^d y_{t-p}\phi_p \\ &\quad +\varepsilon_{t-1}\rho_1 + \varepsilon_{t-2}\rho_2 + \ldots + \varepsilon_{t-q}\rho_q \\ &\quad +\mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t\end{aligned}$$

This just an ARMA(p,q) model applied to differenced y_t

The same MLE that gave us ARMA estimates still estimates $\hat{\boldsymbol{\phi}}$, $\hat{\boldsymbol{\rho}}$, and $\hat{\boldsymbol{\beta}}$

We just need to choose d based on theory, ACFs and PACFs, and unit root tests (ugh)

ARIMA(p,d,q) models

Conditional forecasting and in-sample counterfactuals work just as before

Same code from last time will work; just change the d term of the ARIMA order to 1

Example: Presidential Approval

We have data on the percent ($\times$ 100) of Americans supporting President Bush, averaged by month, over 2/2001–6/2006.

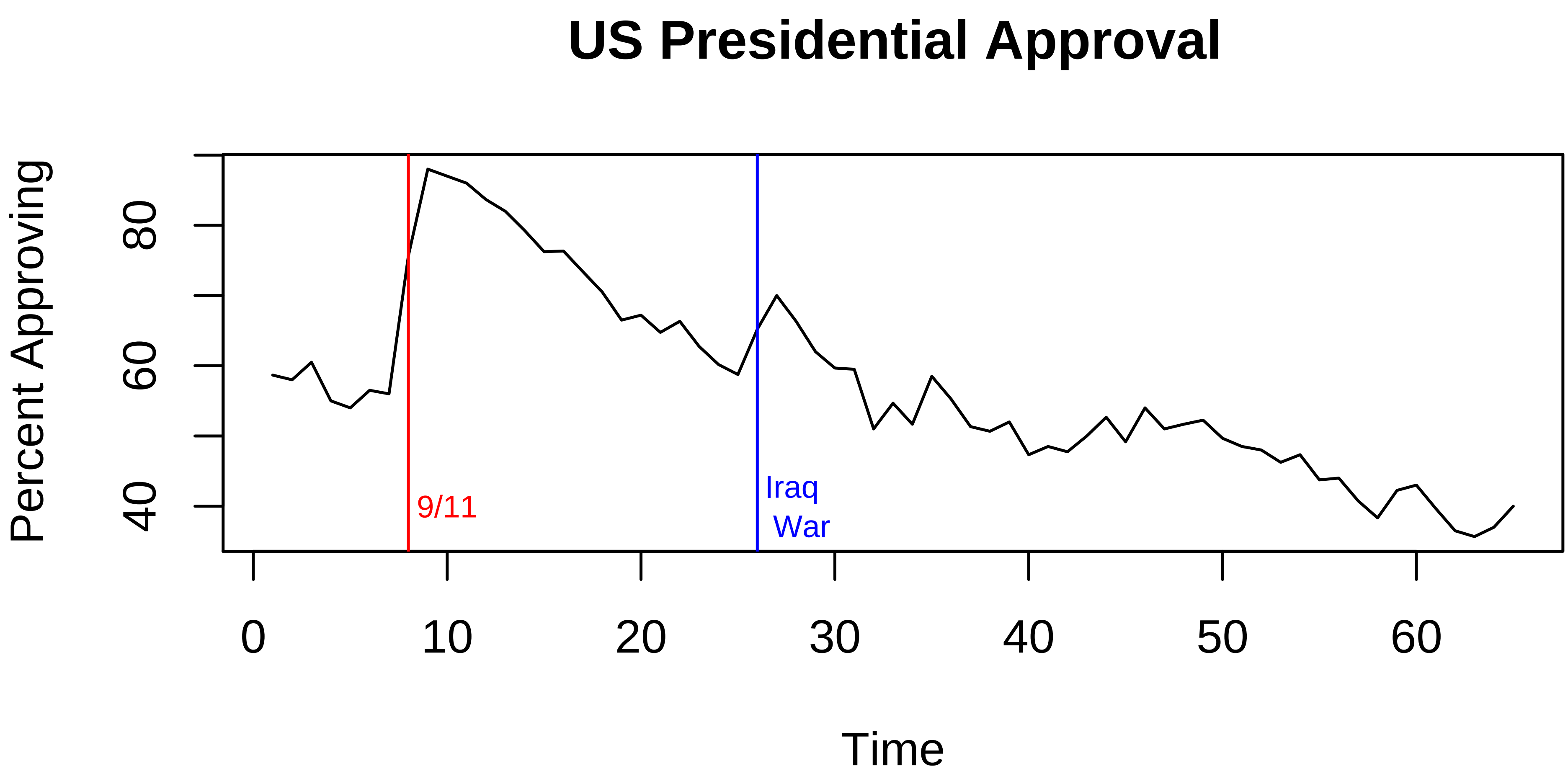
Our covariates include:

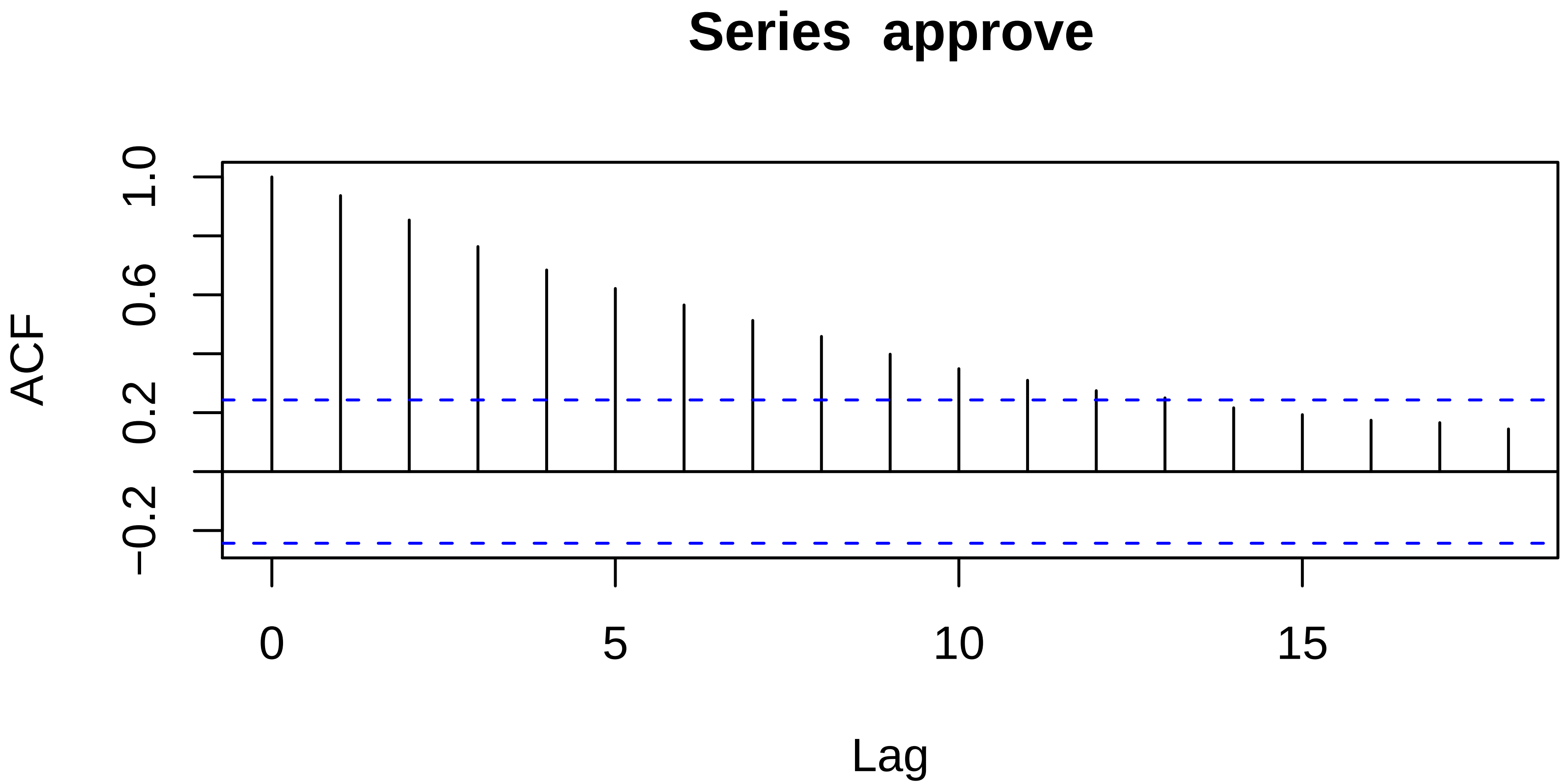
The average price of oil per month, in $/barrel

Dummies for September and October of 2001

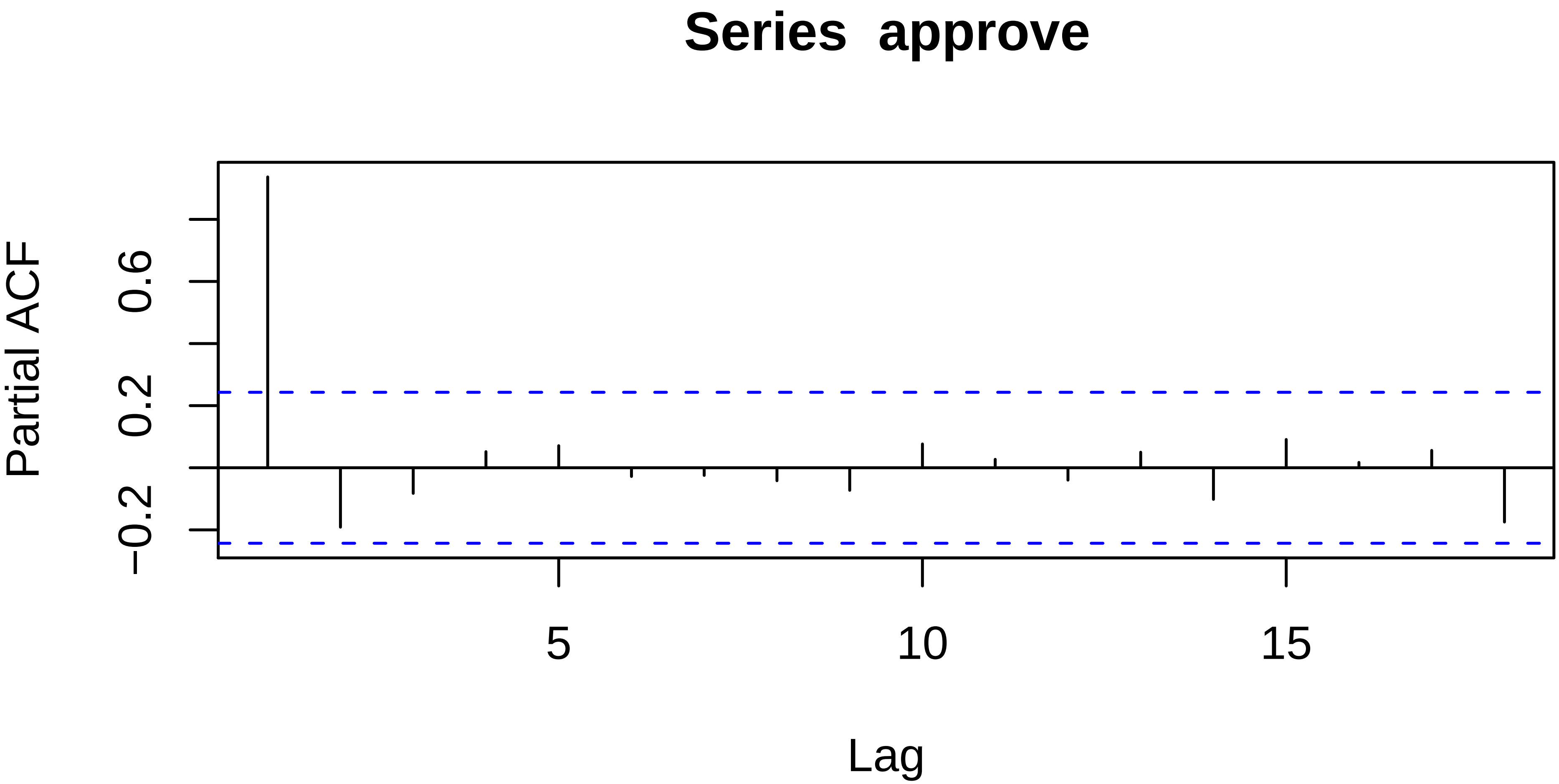
Dummies for first three months of the Iraq War

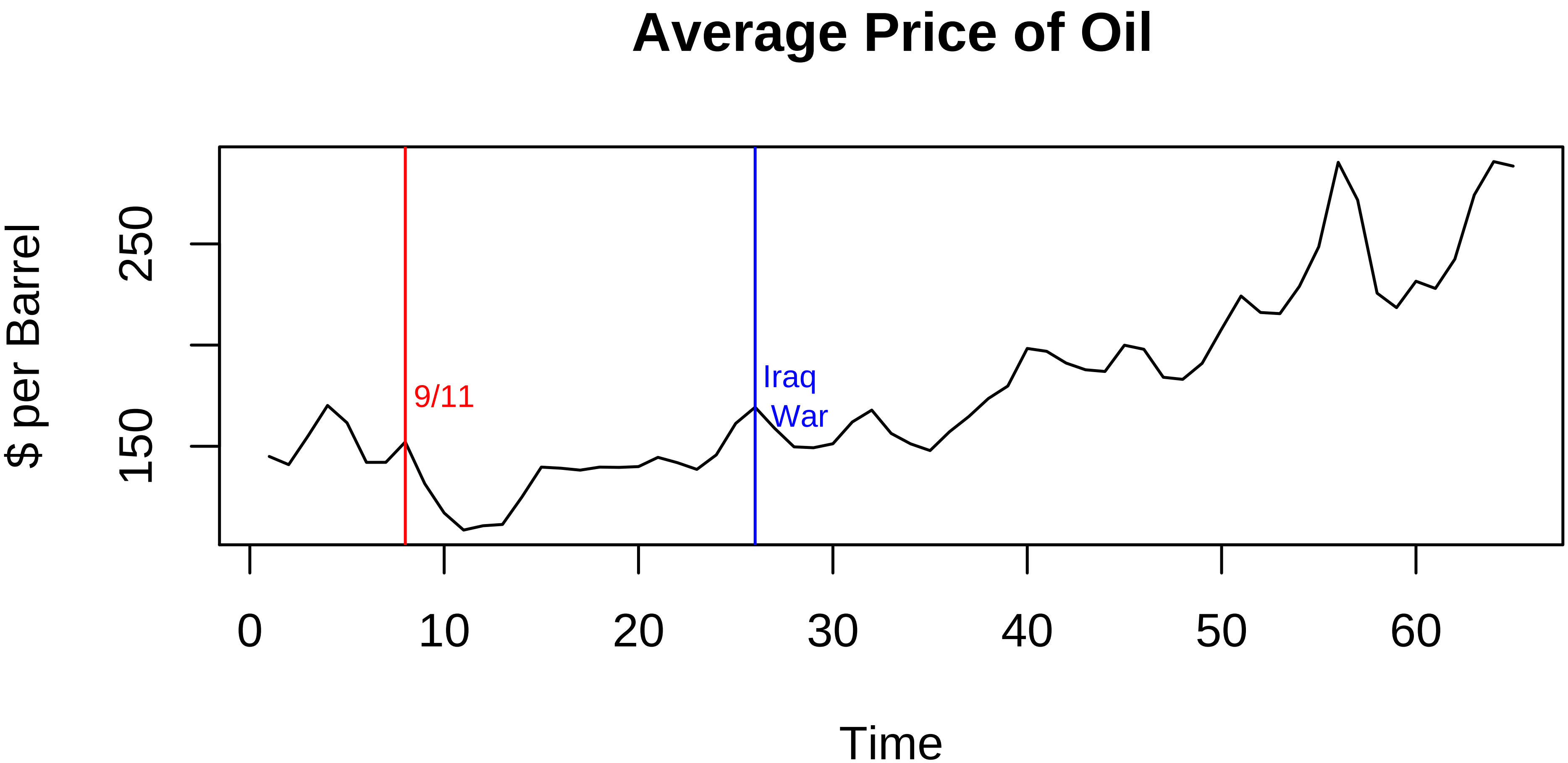
Let's look at our two continuous time series


US Presidential Approval
Percent Approving
40
60
80
9/11
Iraq
War
0
10
20
30
40
50
60
Time

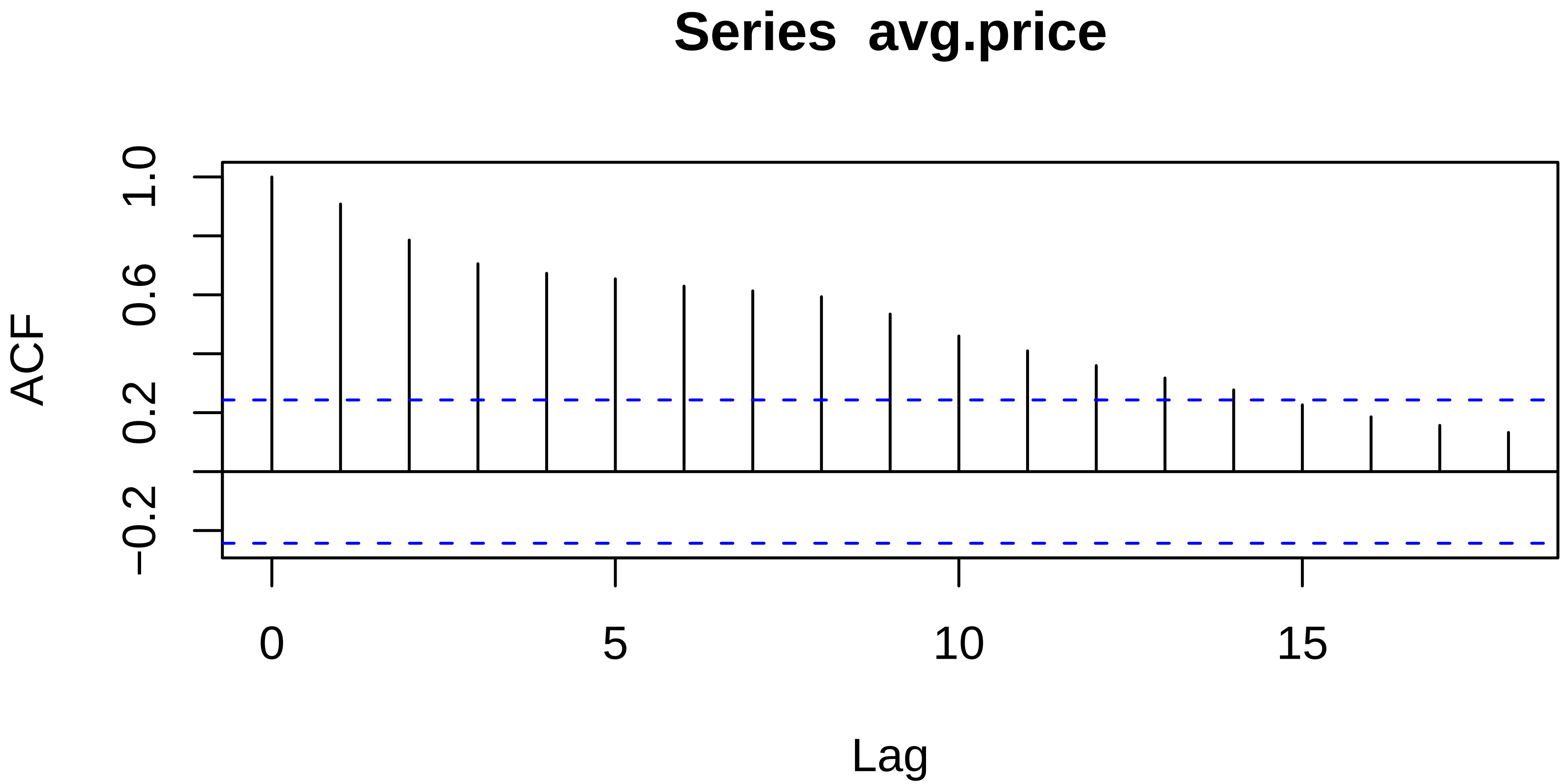


Series approve
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag

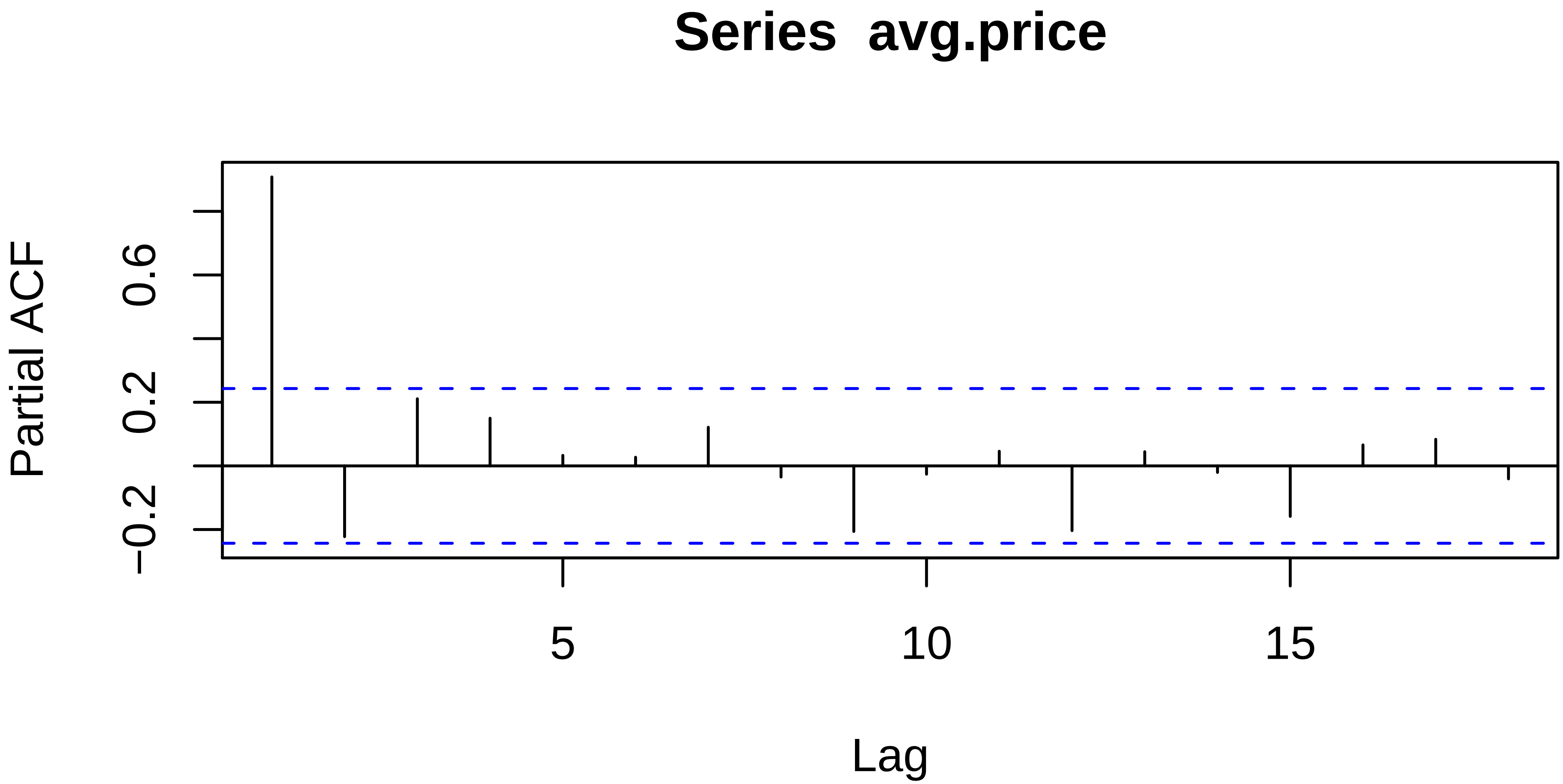

Series approve
Partial ACF
-0.2
0.2
0.6
5
10
15
Lag



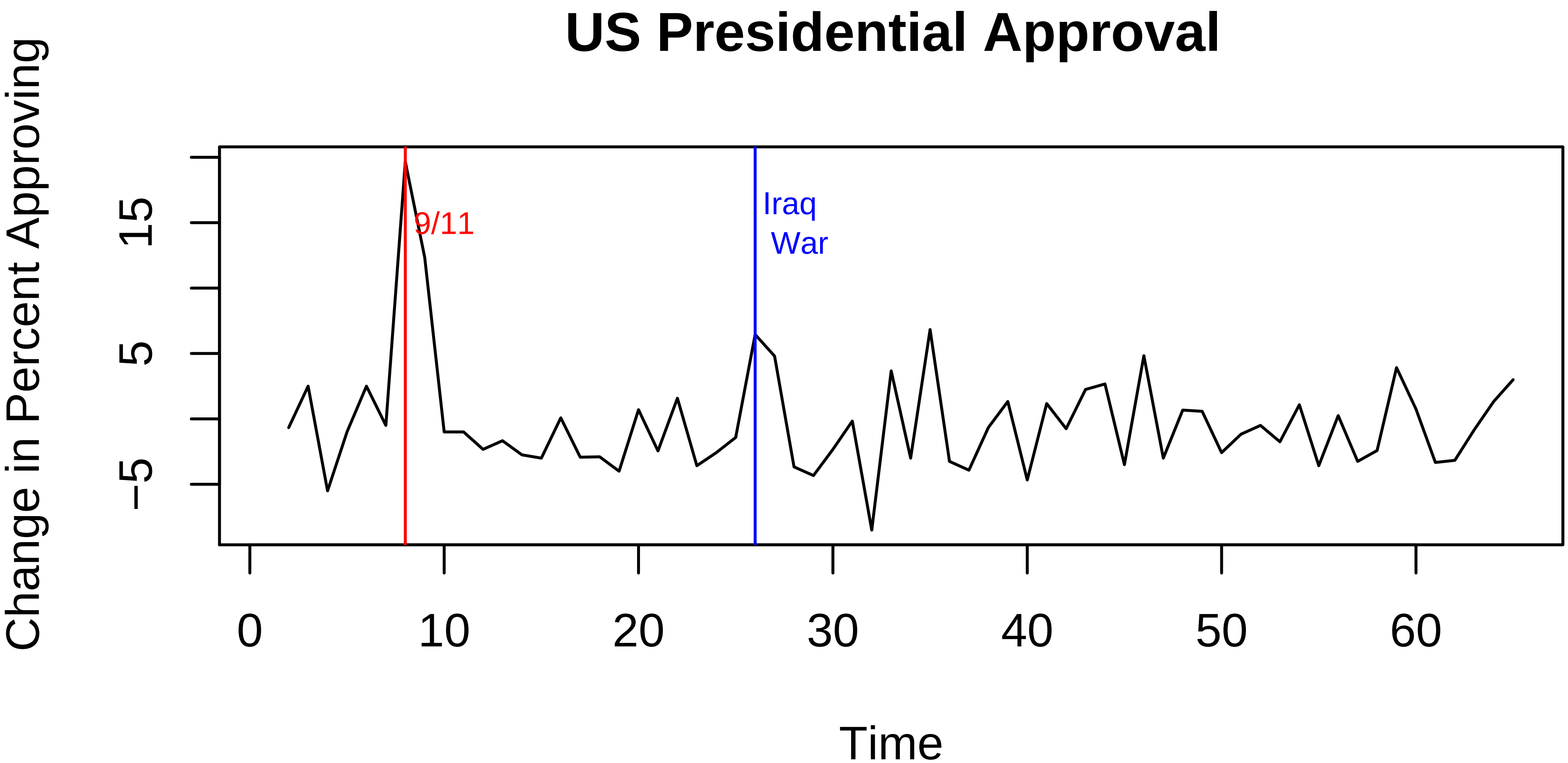
Average Price of Oil
$ per Barrel
150
250
9/11
Iraq
War
0
10
20
30
40
50
60
Time

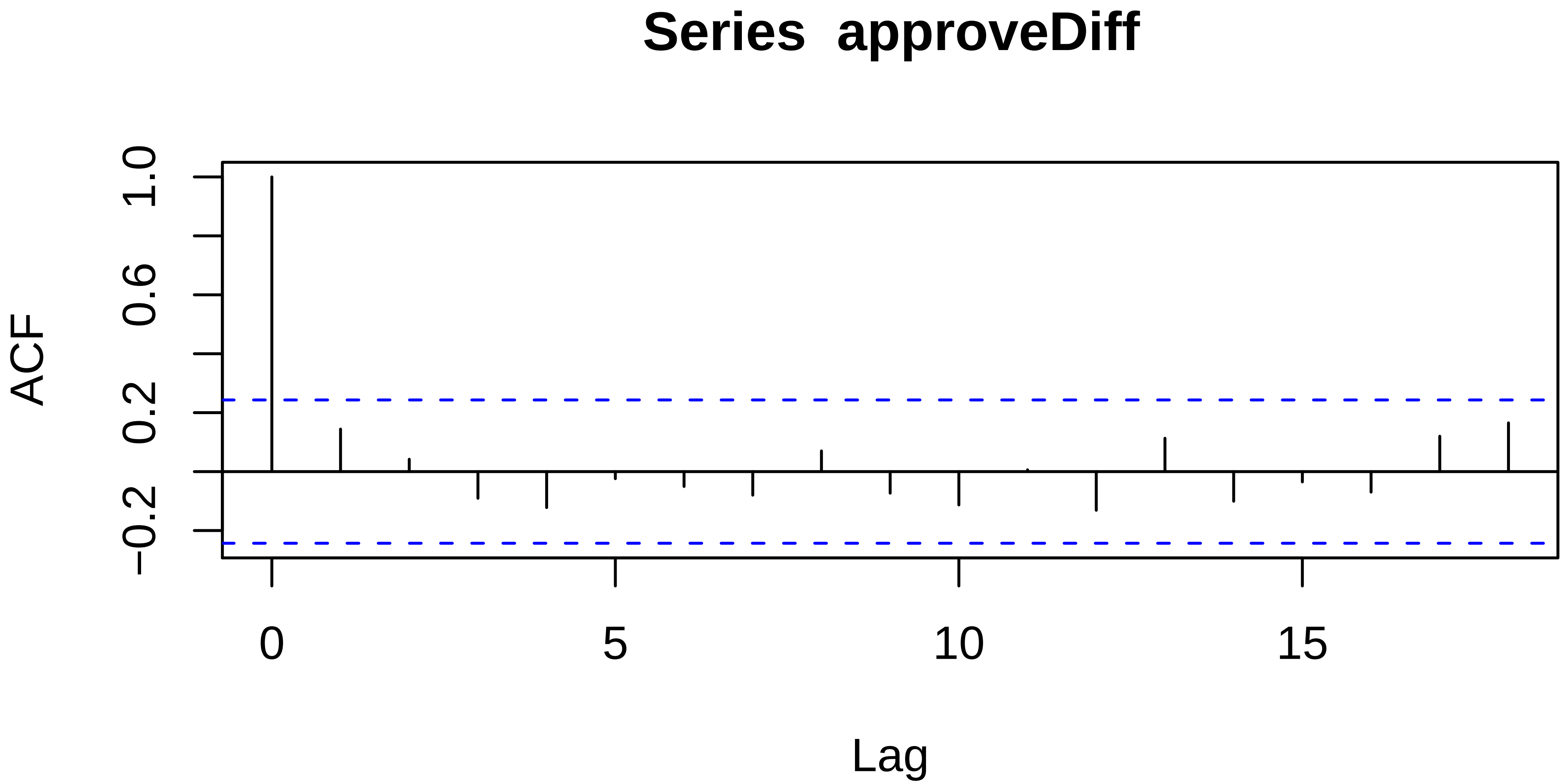


Series avg.price
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag

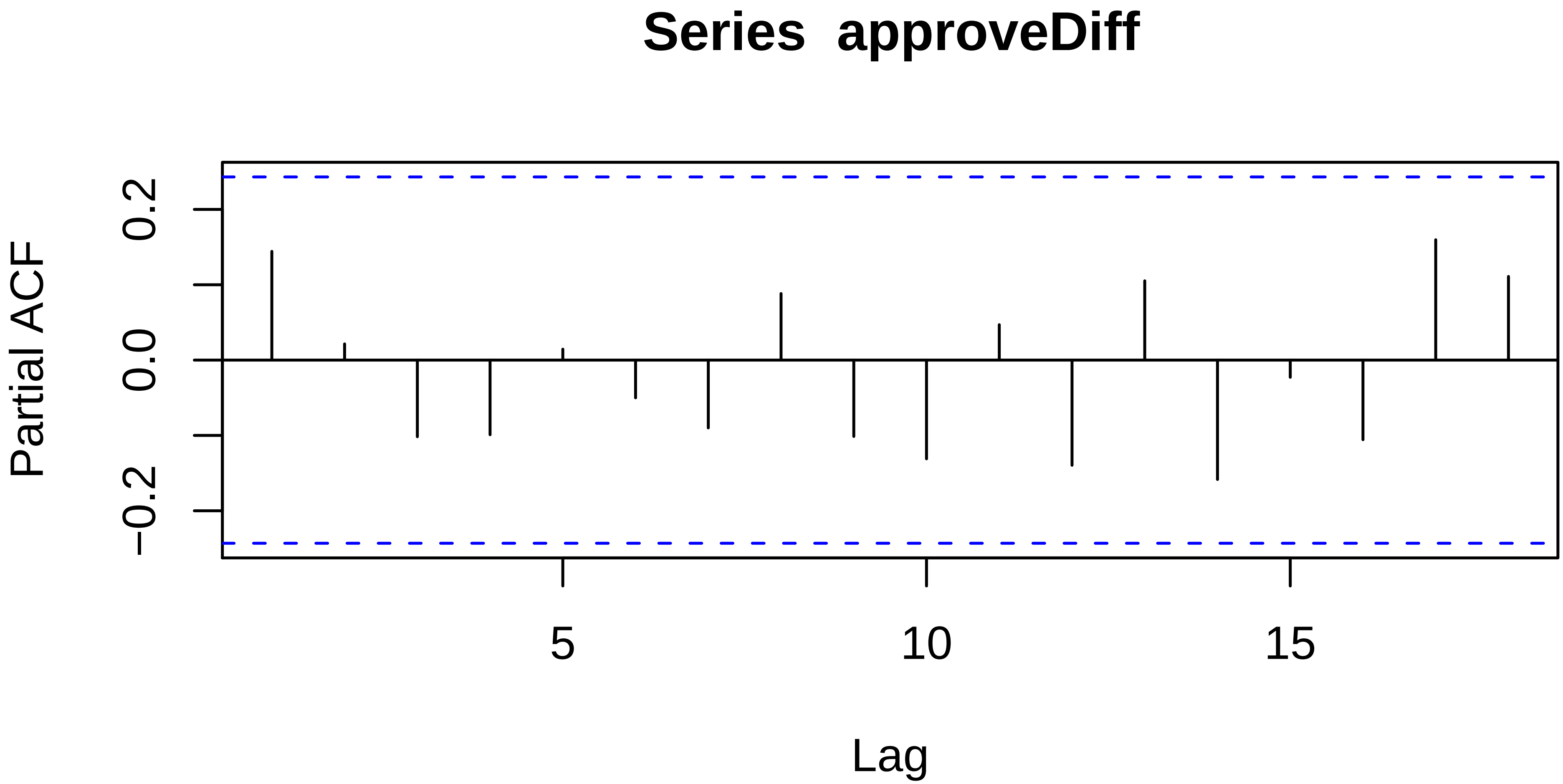


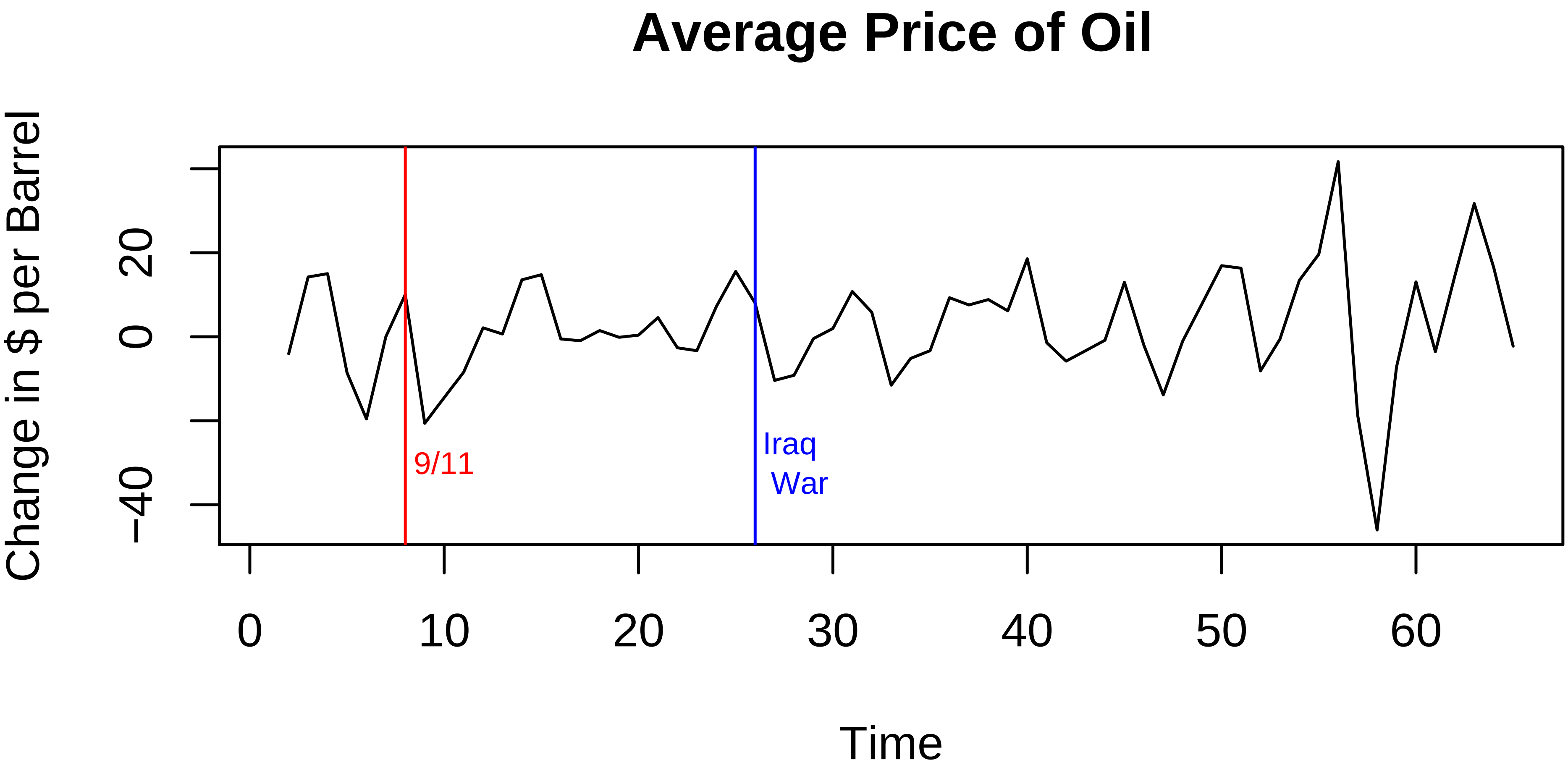
Series avg.price
Partial ACF
-0.2
0.2
0.6
5
10
15
Lag


US Presidential Approval
Change in Percent Approving
−5
5
15
9/11
Iraq
War
0
10
20
30
40
50
60
Time

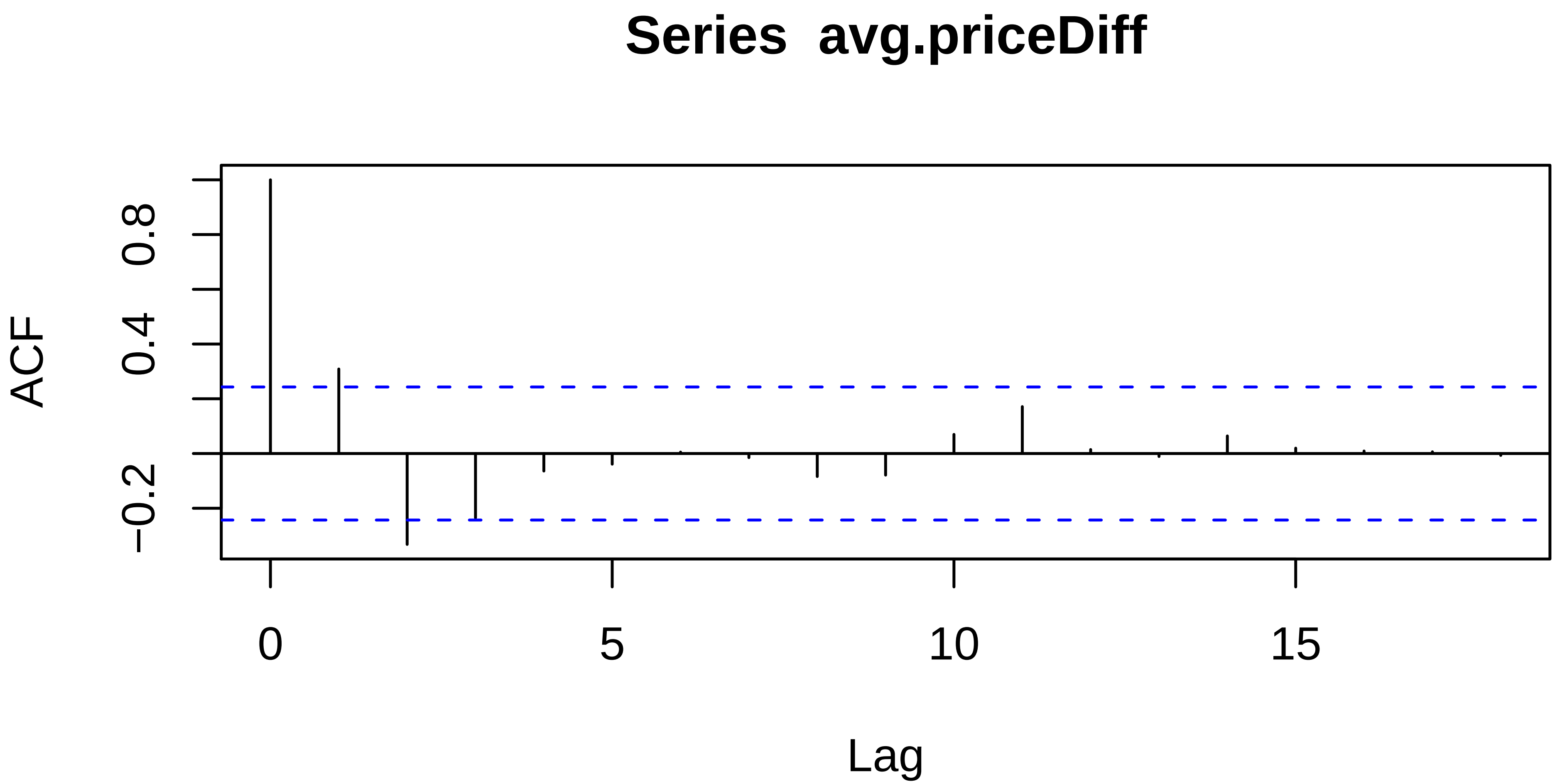


Series approveDiff
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag

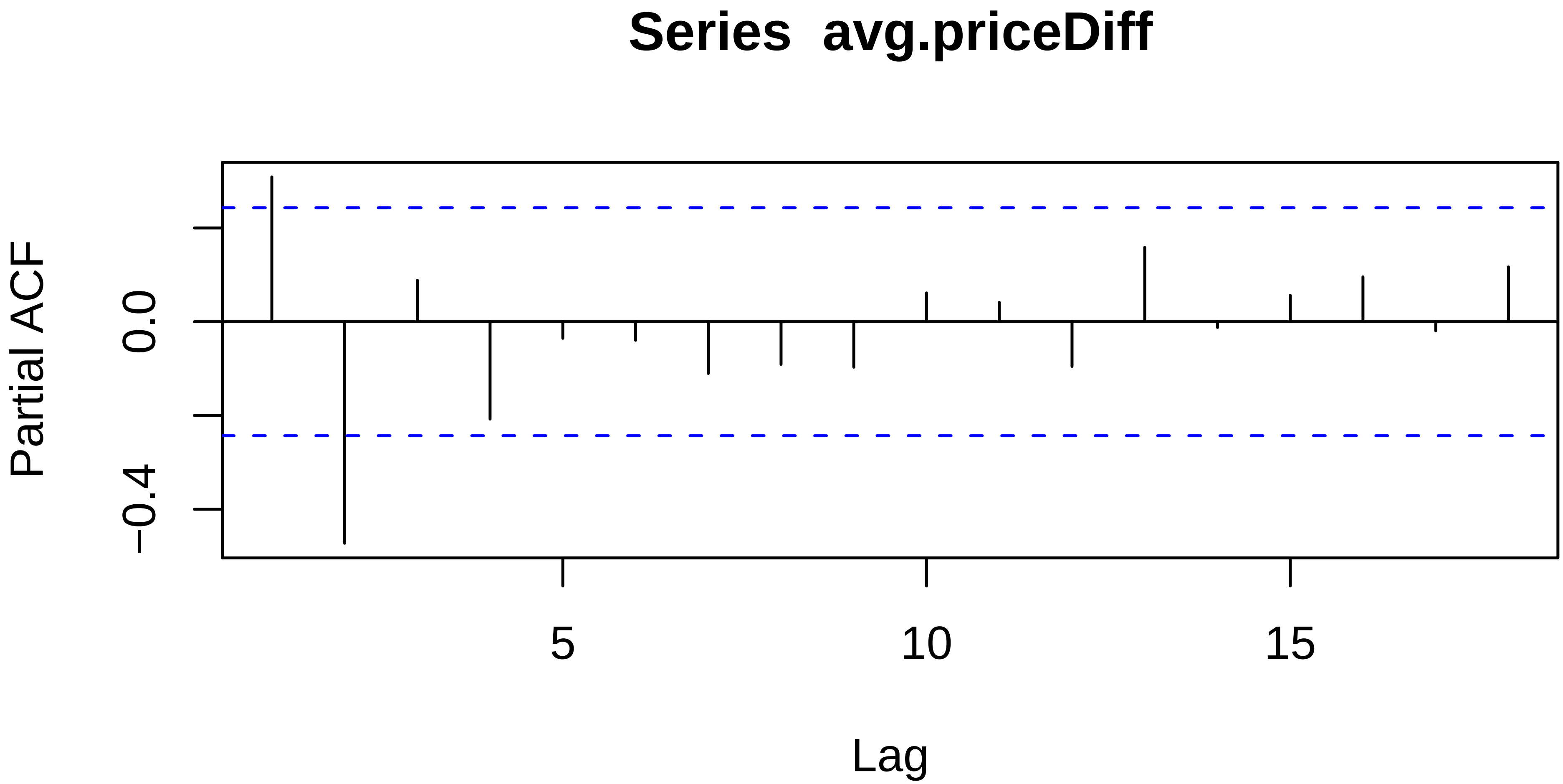

Series approveDiff
Partial ACF
0.2
0.0
−0.2
5
10
15
Lag



Average Price of Oil
Change in $ per Barrel
−40
0
20
9/11
Iraq
War
0
10
20
30
40
50
60
Time



Series avg.priceDiff
ACF
0.8
0.4
−0.2
0
5
10
15
Lag


Series avg.priceDiff
Partial ACF
0.0
−0.4
5
10
15
Lag

Example: Presidential Approval

Many suspect `approve` and `avg.price` are non-stationary processes

Theoretically, what does this mean? Could an approval rate drift anywhere?

Note a better dependent variable would be the logit transformation of approve, $\ln(\texttt{approve}/(1 - \texttt{approve}))$, which is unbounded and probably closer to the latent concept of support

And extenting `approve` out to $T = \infty$ would likely stretch the concept too far for a democracy with regular, anticipated elections

We'll ignore this to focus on the TS issues

Example: Presidential Approval

To a first approximation, we suspect `approve` and `avg.price` may be non-stationary processes

We know that regressing one I(1) process on another risks spurrious correlation

How can we investigate the relationship between these variables?

Strategy 1: ARIMA(0,1,0), first differencing

Example: Presidential Approval

We load the data, plot it, with ACFs and PACFs

Then perform unit root tests

```
> PP.test(approve)

        Phillips-Perron Unit Root Test

data:  approve
Dickey-Fuller = -2.839, Truncation lag parameter = 3, p-value = 0.2350

> adf.test(approve)

        Augmented Dickey-Fuller Test

data:  approve
Dickey-Fuller = -3.957, Lag order = 3, p-value = 0.01721
alternative hypothesis: stationary
```

Example: Presidential Approval

```
> PP.test(avg.price)

        Phillips-Perron Unit Root Test

data:  avg.price
Dickey-Fuller = -2.332, Truncation lag parameter = 3, p-value = 0.4405

> adf.test(avg.price)

        Augmented Dickey-Fuller Test

data:  avg.price
Dickey-Fuller = -3.011, Lag order = 3, p-value = 0.1649
alternative hypothesis: stationary
```

Example: Presidential Approval

We create differenced versions of the time series, and repeat

```
> adf.test(na.omit(approveDiff))

        Augmented Dickey-Fuller Test

data:  na.omit(approveDiff)
Dickey-Fuller = -4.346, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary

> adf.test(na.omit(avg.priceDiff))

        Augmented Dickey-Fuller Test

data:  na.omit(avg.priceDiff)
Dickey-Fuller = -5.336, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

Example: Presidential Approval

We estimate an ARIMA(0,1,0), which fit a little better than ARIMA(2,1,2) on the AIC criterion

```
Call:
arima(x = approve, order = c(0, 1, 0),
      xreg = xcovariates, include.mean = TRUE)

Coefficients:
      sept.oct.2001  iraq.war  avg.price
             11.207     5.690     -0.071
s.e.          2.519     2.489      0.034

sigma^2 estimated as 12.4:  log likelihood = -171.2,  aic = 350.5
```

Example: Presidential Approval

To interpret the model, we focus on historical counterfactuals

What would Bush's approval have looked like if 9/11 hadn't happened?

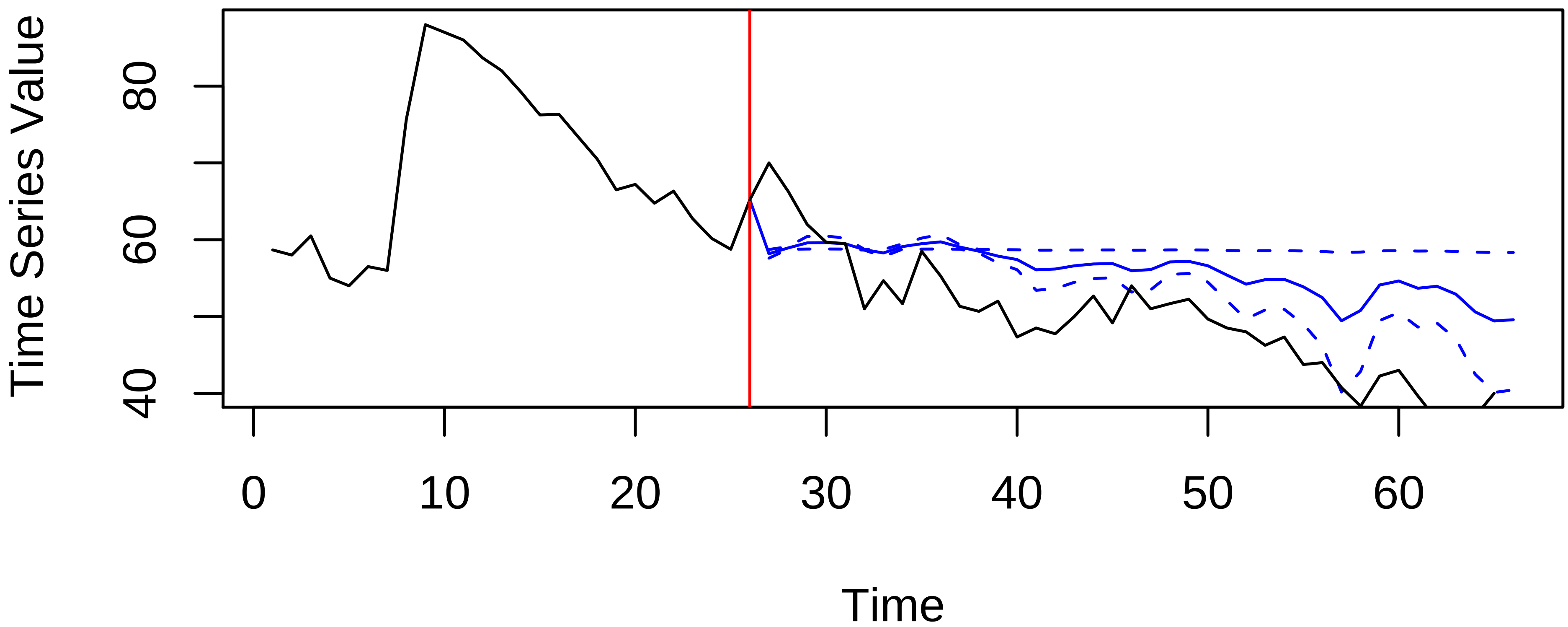
What if Bush had not invaded Iraq?

What if the price of oil had remained at pre-war levels?

Naturally, we only trust our results so far as we trust the model

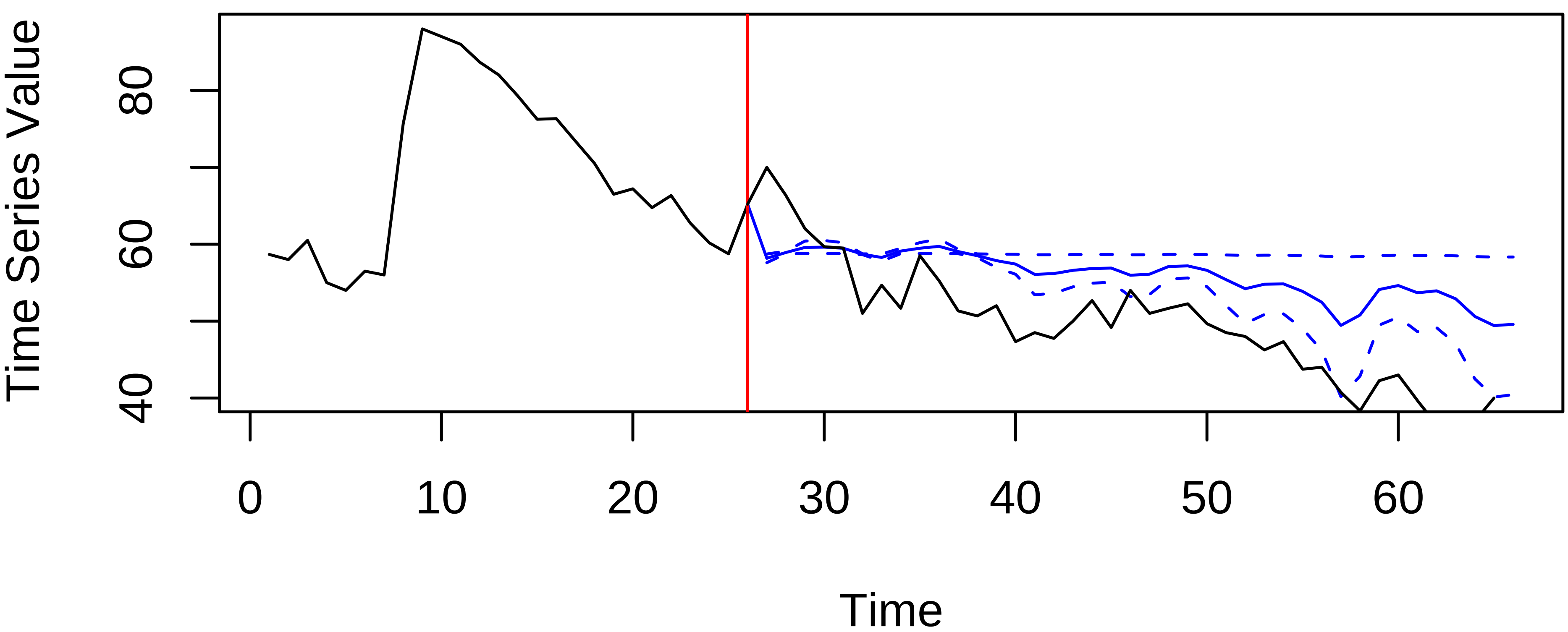
(which is not very much—we've left out a lot, like unemployment, inflation, boundedness of approve, . . .)

We simulate counterfactual approval using Zelig's implementation of ARIMA



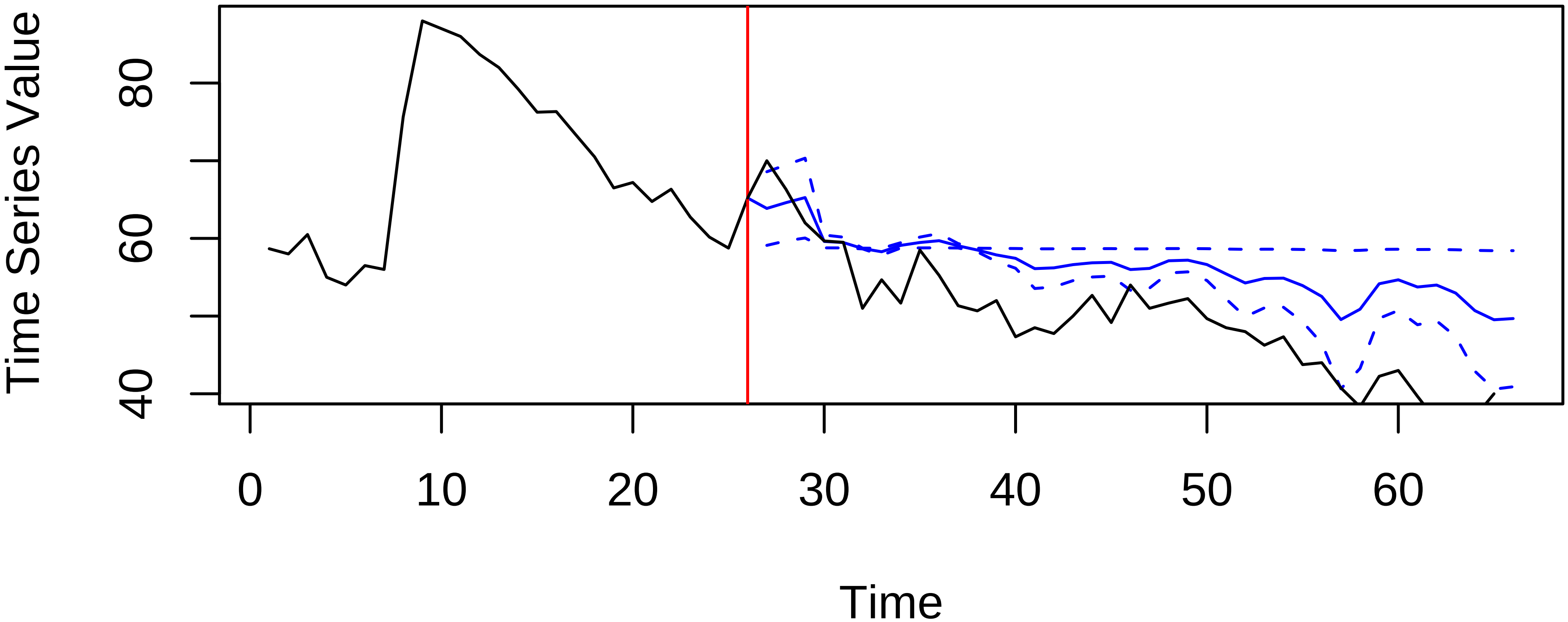
In blue: Predicted Bush approval without Iraq

In black: Actual approval



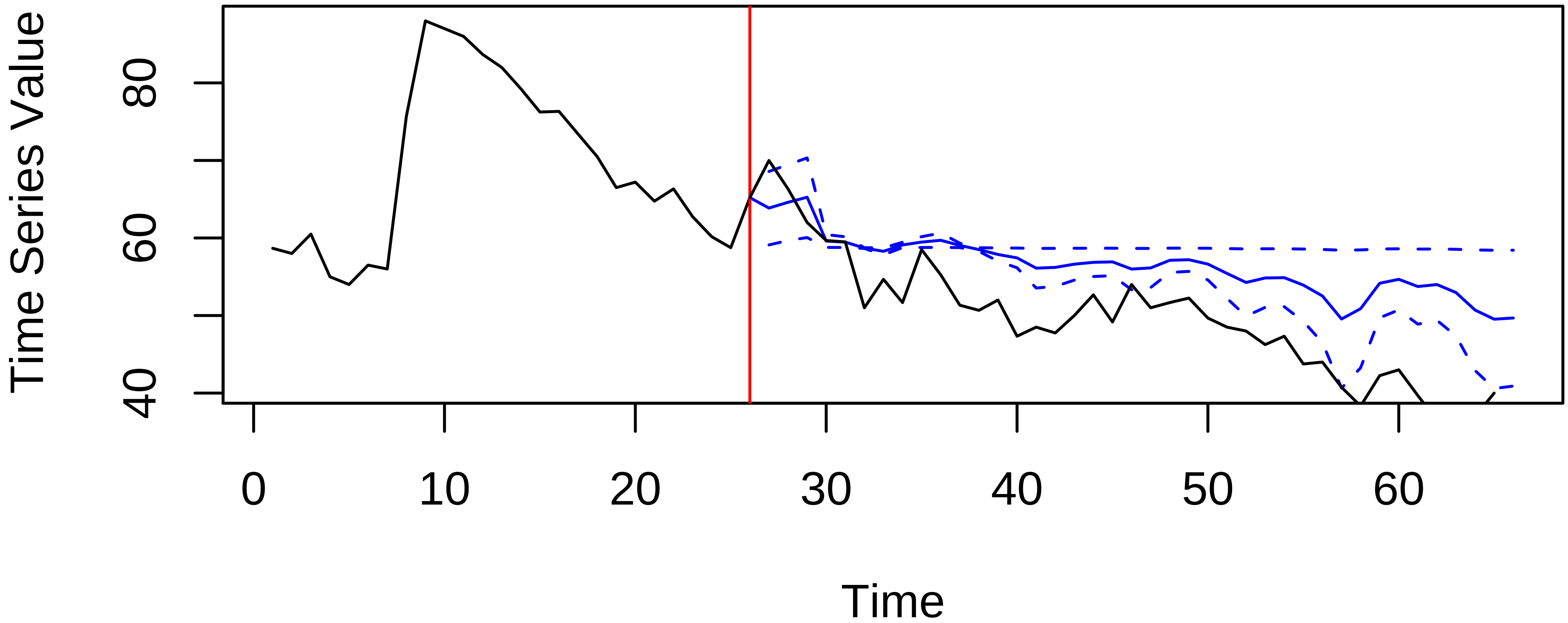
At first, starting the war in Iraq appears to help Bush's popularity

Then, it hurts—a lot. Sensible result. So are we done?

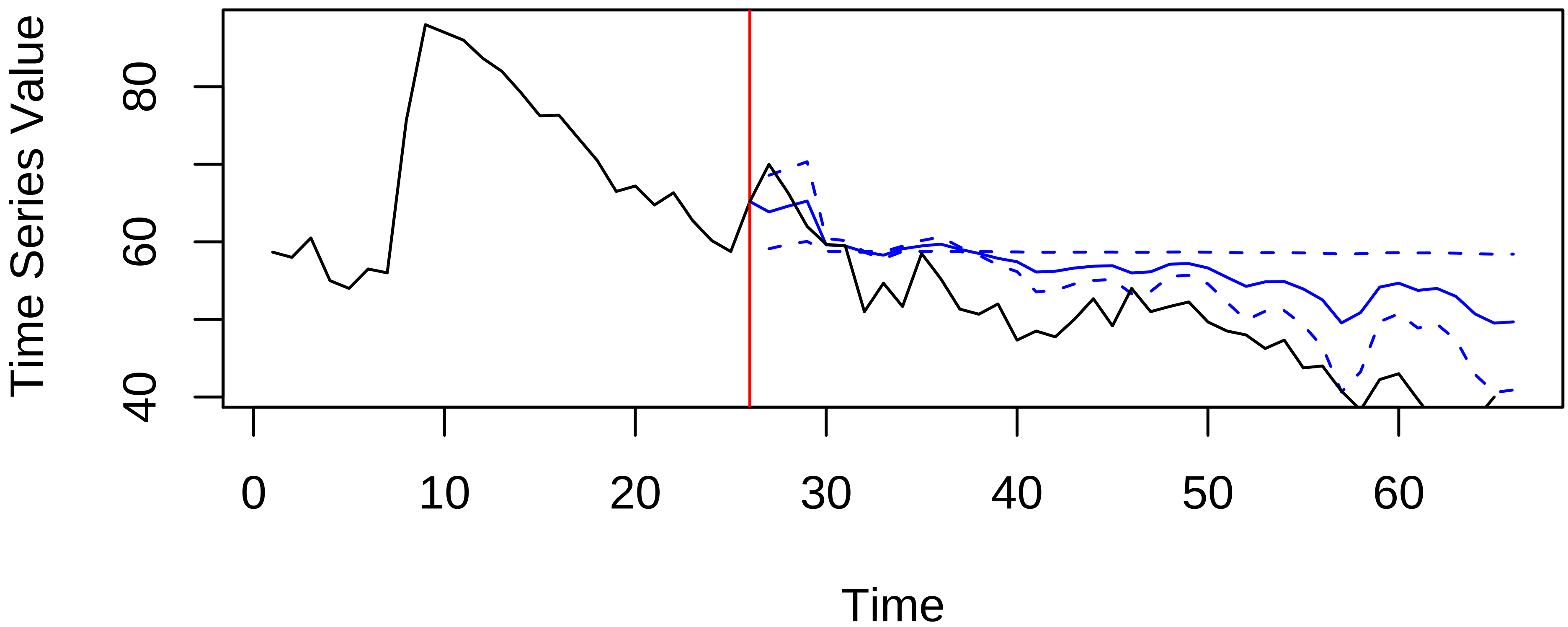


In blue: Predicted Bush approval with Iraq war

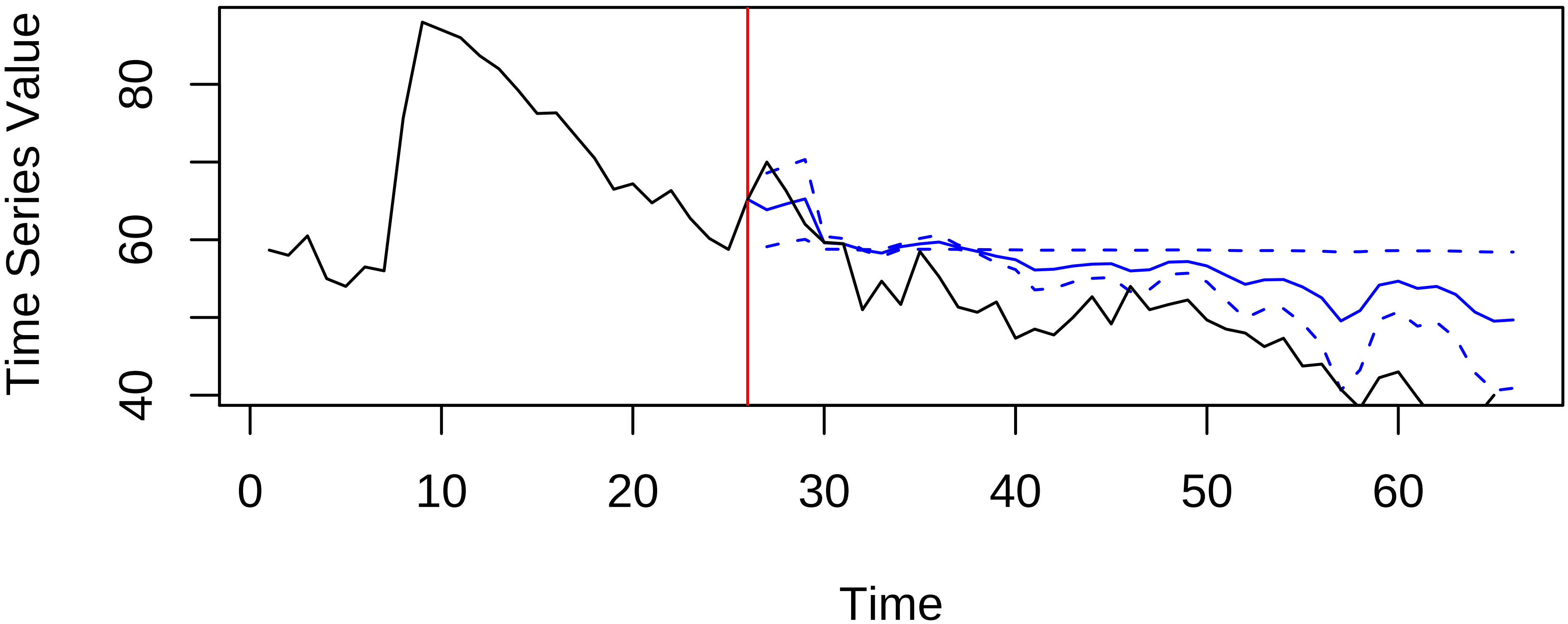
In black: Actual approval



Wait—can the model predict the long run approval rate?

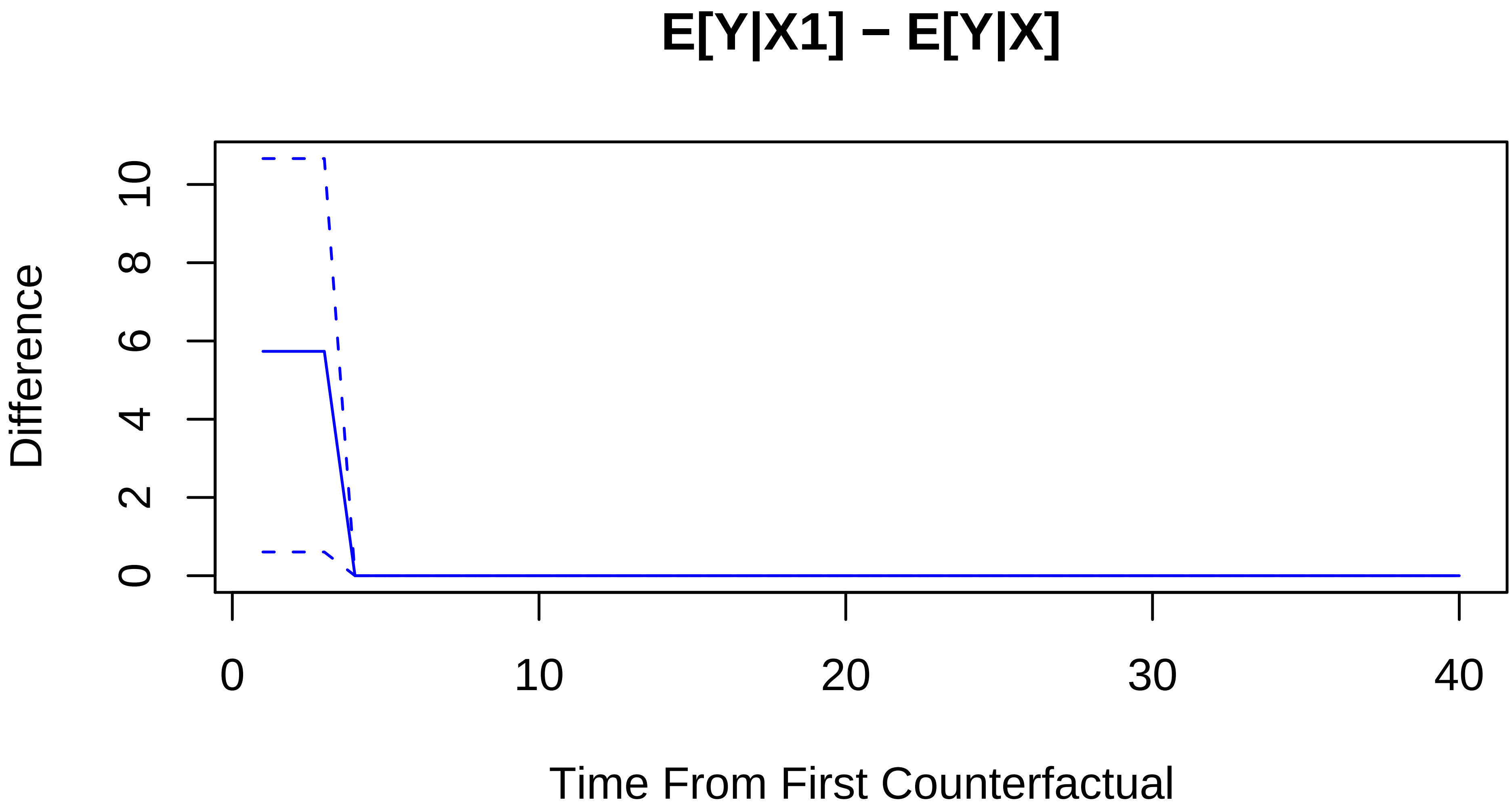


Wait—can the model predict the long run approval rate? Not even close



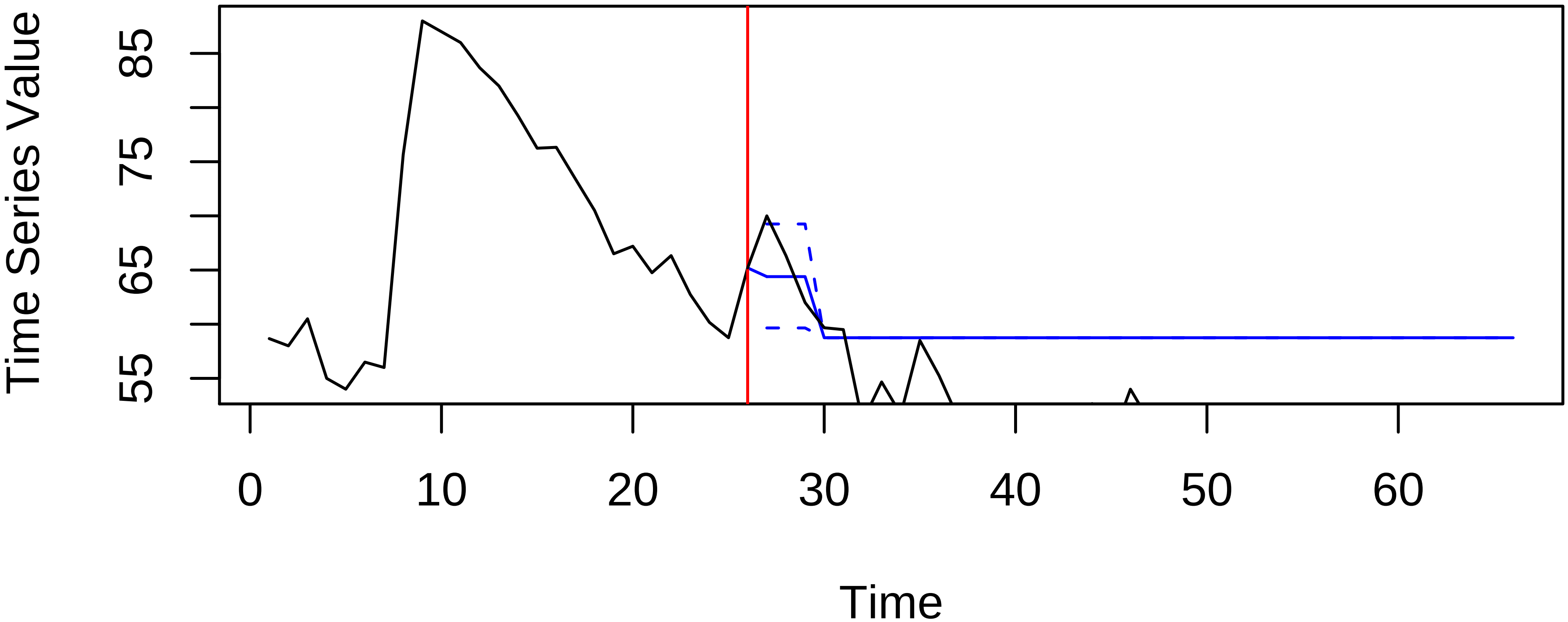
The model fit well for the first few months, then stays close to the ex ante "mean" approval

But reality (which is I(1)) drifts off into the cellar



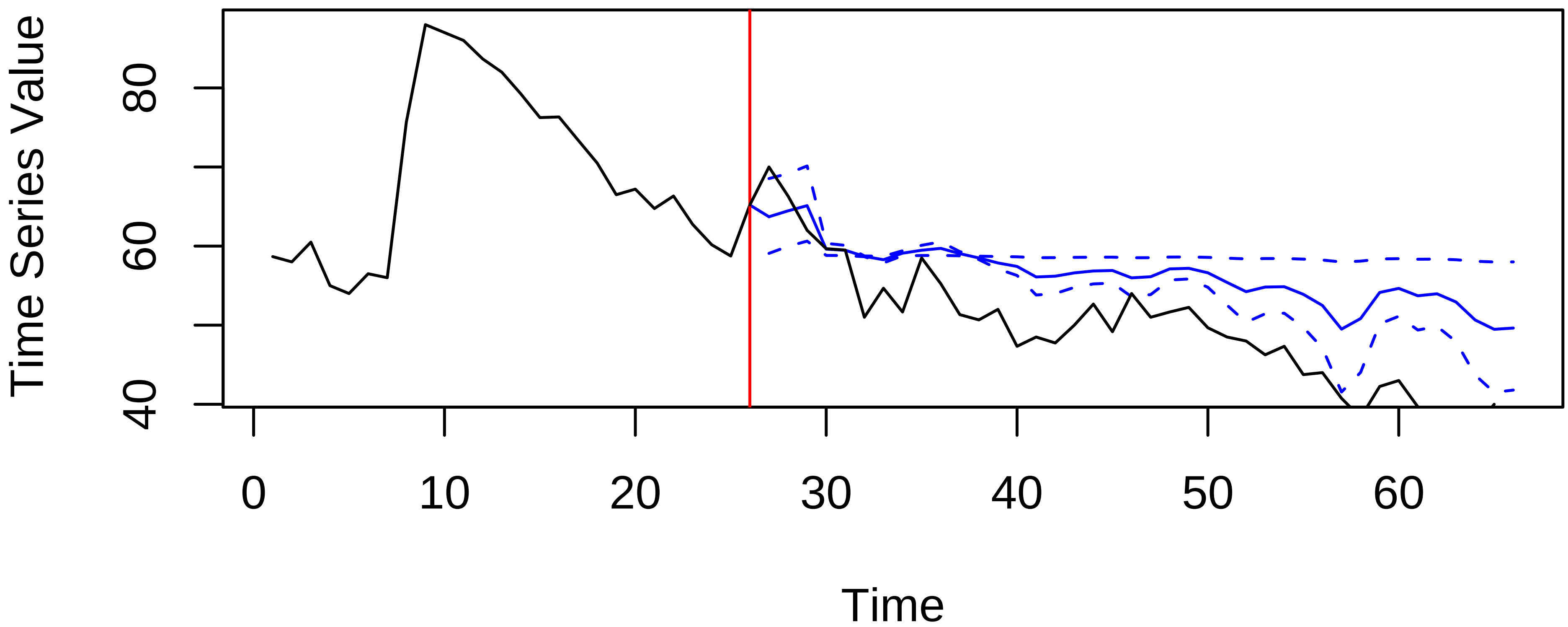
First differences show that all the action is in the short-run

Long-run predictions are not feasible with unit root processes



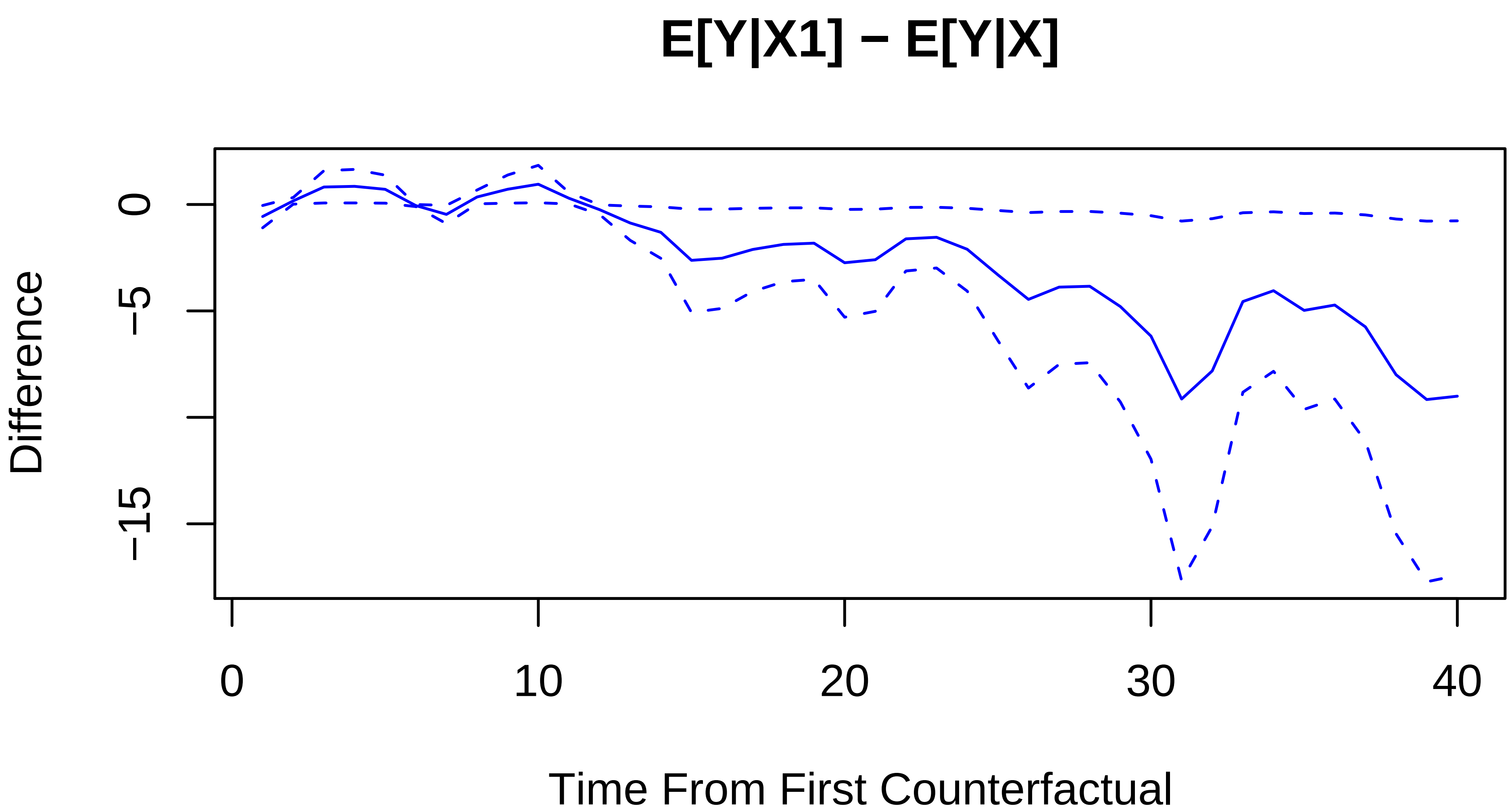
Suppose Oil had stayed at its pre-war price of $161/barrel

Then Bush's predicted popularity looks higher than the data



But wait—here are the factual “predictions” under the actual oil price

Miss the data by a mile


E[Y|X1] − E[Y|X]
Difference
0
−5
−15
0
10
20
30
40
Time From First Counterfactual

Limits of ARIMA

ARIMA(p,1,q) does a good job of estimating the short run movement of stationary variables

But does a terrible job with long-run levels

No surprise: The model includes no level information

While the observed level could drift anywhere

Limits of ARIMA

Using Δy_t as our response has a big cost

Purging all long-run equilibrium relationships from our time series

These empirical long-run relationships may be spurious (why we're removing them)

But what if they are not? What if y_t and x_t really move together over time?

Then removing that long-run relationship removes theoretically interesting information from our data

Since most of our theories are about long-run levels of our variables, we have usually just removed the *most* interesting part of our dataset!

Cointegration

Consider two time series y_t and x_t:

$$
\begin{aligned}
x_t &= x_{t-1} + \varepsilon_t \\
y_t &= y_{t-1} + 0.6x_t + \nu_t
\end{aligned}
$$

where ε_t and ν_t are (uncorrelated) white noise

x_t and y_t are both: AR(1) processes, random walks, non-stationary, and I(1).

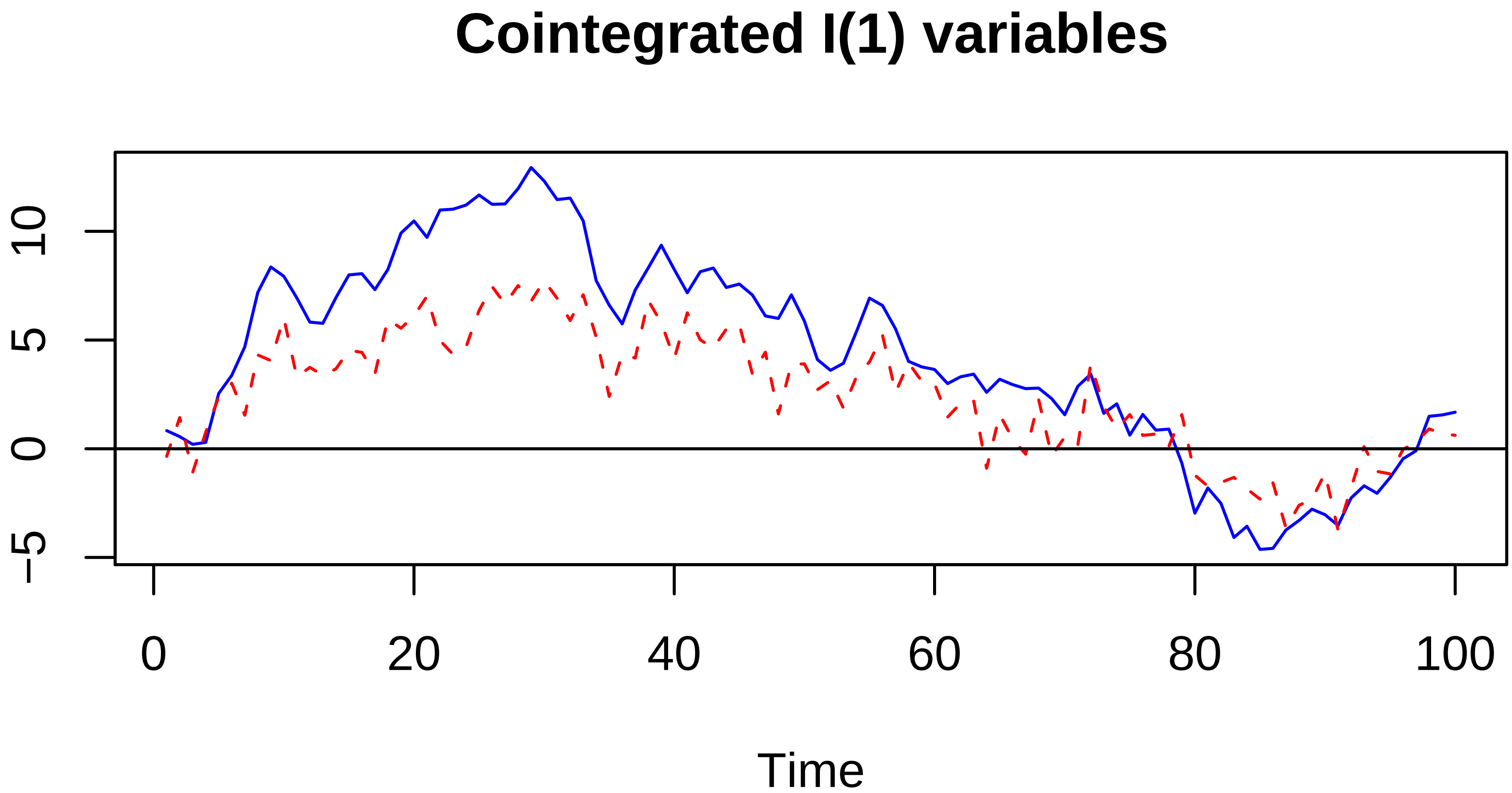
They are not spuriously correlated, but genuinely causally connected

Neither tends towards any particular level, but each tends towards the other

A particularly large ν_t may move y_t away from x_t briefly,
but eventually, y_t will move back to x_t's level

As a result, they will move together through t indefinitely

x_t and y_t are said to be *cointegrated*


Cointegrated I(1) variables
10
5
0
−5
0
20
40
60
80
100
Time

Cointegration

Any two (or more) variables y_t, x_t are said to be cointegrated if

1. each of the variables is I(1)

2. there is some vector $\boldsymbol{\alpha}$ such that

$$
\begin{aligned}
z_t &= \mathrm{cbind}(y_t, x_t)\boldsymbol{\alpha} \\
z_t &\sim I(0)
\end{aligned}
$$

   or in words, there is some linear combination of the non-stationary variables which is stationary

There may be many cointegrating vectors;
the cointegration rank r gives their number

Cointegration: Engle-Granger Two Step

Several ways to find the cointegration vector(s) and use it to anaylze the system

Simplest is Engle-Granger Two Step Method

Works best if cointegration rank is $r = 1$

Cointegration: Engle-Granger Two Step

Several ways to find the cointegration vector(s) and use it to anaylze the system

Simplest is Engle-Granger Two Step Method

Works best if cointegration rank is $r = 1$

Step 1: Estimate the cointegration vector by least squares with no constant:

$$y_t = \alpha_1^* x_{t-1} + \alpha_2^* x_{t-2} + \ldots + \alpha_K^* x_{t-K} + z_t$$

This gives us the cointegrating vector $\alpha = (1, -\alpha_1^*, -\alpha_2^*, \ldots - \alpha_K^*)$

and the long-run equilibrium path of the cointegrated variables, $\hat{z}_t$

We can test for cointegration by checking that $\hat{z}_t$ is stationary

Note that the usual unit root tests work, but with different critical values

This is because the $\hat{\alpha}$'s are very well estimated: "super-consistent"
(converge to their true values very fast as T increases)

Cointegration: Engle-Granger Two Step

Step 2: Estimate an Error Correction Model

After obtaining the cointegration $\hat{z}_t$ and confirming it is I(0), we can estimate a particularly useful specification known as an *error correction model*, or ECM

ECMs simultaneously estimate long- and short-run effects for a system of cointegrated variables

Better than ARI(p,d) because we don't throw away level information

Interestingly, can be estimated with least squares

Cointegration: Engle-Granger Two Step

For a bivariate system of y_t, x_t, two equations describe how this cointegrated process evolves over time:

$$\Delta y_t \quad = \quad \psi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{j=1}^{J} \psi_{1j} \Delta x_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta y_{t-k} + u_t$$

Cointegration: Engle-Granger Two Step

For a bivariate system of y_t, x_t, two equations describe how this cointegrated process evolves over time:

$$\Delta y_t = \psi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{j=1}^{J} \psi_{1j} \Delta x_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta y_{t-k} + u_t$$

$$\Delta x_t = \zeta_0 + \gamma_2 \hat{z}_{t-1} + \sum_{j=1}^{J} \zeta_{1j} \Delta y_{t-j} + \sum_{k=1}^{K} \zeta_{2k} \Delta x_{t-k} + v_t$$

Cointegration: Engle-Granger Two Step

For a bivariate system of y_t, x_t, two equations describe how this cointegrated process evolves over time:

$$\begin{aligned}
\Delta y_t &= \psi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{j=1}^{J} \psi_{1j} \Delta x_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta y_{t-k} + u_t \\
\Delta x_t &= \zeta_0 + \gamma_2 \hat{z}_{t-1} + \sum_{j=1}^{J} \zeta_{1j} \Delta y_{t-j} + \sum_{k=1}^{K} \zeta_{2k} \Delta x_{t-k} + v_t
\end{aligned}$$

These equations are the "error correction" form of the model

Show how y_t and x_t respond to deviations from their long run relationship

Cointegration: Engle-Granger Two Step

Let's focus on the evolution of y_t as a function of its lags, lags of x_t, and the error in the long-run equilibrium, $\hat{z}_{t-1}$:

$$\Delta y_t = \psi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{j=1}^{J} \psi_{1j} \Delta x_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta y_{t-k} + u_t$$

Cointegration: Engle-Granger Two Step

Let's focus on the evolution of y_t as a function of its lags, lags of x_t, and the error in the long-run equilibrium, $\hat{z}_{t-1}$:

$$\Delta y_t \quad = \quad \psi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{j=1}^{J} \psi_{1j} \Delta x_{t-j} + \sum_{k=1}^{K} \psi_{2k} \Delta y_{t-k} + u_t$$

$\gamma_1 < 0$ must hold: This is the speed of adjustment back to equilibrium; larger negative values imply faster adjustment

This is the central assumption of cointegration:
In the long run, y_t and x_t cannot diverge

So short-run differences must be made up later by convergence

For example, y_t must *eventually reverse course* after a big shift away from x_t

γ_1 shows how quickly y_t reverse back to x_t

Cointegration: Engle-Granger Two Step

Recall our cointegrated time series, y_t and x_t:

$$
\begin{aligned}
x_t &= x_{t-1} + \varepsilon_t \\
y_t &= y_{t-1} + 0.6 x_t + \nu_t
\end{aligned}
$$

To estimate the Engle-Granger Two Step for these data, we do the following in R:

```
set.seed(123456)

# Generate cointegrated data
e1 <- rnorm(100)
e2 <- rnorm(100)
x <- cumsum(e1)
y <- 0.6*x + e2

# Run step 1 of the E-G two step
coint.reg <- lm(y ~ x)
coint.err <- residuals(coint.reg)
```

```
# Make the lag of the cointegration error term
coint.err.lag <- coint.err[1:(length(coint.err)-2)]

# Make the difference of y and x
dy <- diff(y)
dx <- diff(x)

# And their lags
dy.lag <- dy[1:(length(dy)-1)]
dx.lag <- dx[1:(length(dx)-1)]

# Delete the first dy, because we are missing lags for this obs
dy <- dy[2:length(dy)]

# Estimate an Error Correction Model with LS
ecm1 <- lm(dy ~ coint.err.lag + dy.lag + dx.lag)
summary(ecm1)
```

```
Call:
lm(formula = dy ~ coint.err.lag + dy.lag + dx.lag)

Residuals:
   Min     1Q Median     3Q    Max
-2.959 -0.544  0.137  0.711  2.307

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.0034     0.1036    0.03     0.97
coint.err.lag  -0.9688     0.1585   -6.11  2.2e-08 ***
dy.lag         -1.0589     0.1084   -9.77  5.6e-16 ***
dx.lag          0.8086     0.1120    7.22  1.4e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.03 on 94 degrees of freedom
Multiple R-squared: 0.546,      Adjusted R-squared: 0.532
F-statistic: 37.7 on 3 and 94 DF,  p-value: 4.24e-16
```

Cointegration: Johansen estimator

Alternatively, we can use the urca package, which handles unit roots and cointegration analysis:

```
# Create a matrix of the cointegrated variables
cointvars <- cbind(y,x)

# Perform cointegration tests
coint.test1 <- ca.jo(cointvars,
                   ecdet = "const",
                   type="eigen",
                   K=2,
                   spec="longrun"
                   )
summary(coint.test1)                # Check the cointegration rank her

# Using the output of the test, estimate an ECM
ecm.res1 <- cajorls(coint.test1,
                    r = 1,           # Cointegration rank
                    reg.number = 1)  # which variable(s) to put on LHS
                                     # (column indexes of cointvars)
```

```
summary(ecm.res1$rlm)
```

Cointegration: Johansen estimator

```
######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) , without linear t

Eigenvalues (lambda):
[1]  3.105e-01  2.077e-02 -1.400e-18

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  2.06  7.52  9.24 12.97
r = 0  | 36.44 13.75 15.67 20.20

Eigenvectors, normalised to first column:
(These are the cointegration relations)

             y.l2   x.l2 constant
y.l2      1.00000   1.00    1.000
x.l2     -0.58297  10.13   -1.215
```

```
constant -0.02961 -50.24  -38.501

Weights W:
(This is the loading matrix)

          y.l2      x.l2   constant
y.d -0.967715 -0.001015 -1.004e-18
x.d  0.002461 -0.002817 -2.899e-19
```

Cointegration: Johansen estimator

```
Call:
lm(formula = substitute(form1), data = data.mat)

Residuals:
   Min     1Q Median     3Q    Max
-2.954 -0.536  0.150  0.712  2.318

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
ect1    -0.968      0.158   -6.13  2.0e-08 ***
y.dl1   -1.058      0.108   -9.82  4.1e-16 ***
x.dl1    0.809      0.112    7.26  1.1e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.02 on 95 degrees of freedom
Multiple R-squared: 0.546,      Adjusted R-squared: 0.532
F-statistic: 38.1 on 3 and 95 DF,  p-value: 2.97e-16
```

Example: Approval

Return to our Bush approval example, and estimate an ECM equivalent to the ARIMA(0,1,0) model we chose:

```
Residuals:
   Min     1Q Median     3Q    Max
-7.140 -1.675 -0.226  1.643  5.954

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
ect1           -0.1262     0.0301   -4.20  9.4e-05 ***
sept.oct.2001  19.5585     2.1174    9.24  5.4e-13 ***
iraq.war        5.0187     1.6243    3.09   0.0031 **
approve.dl1    -0.3176     0.0945   -3.36   0.0014 **
avg.price.dl1  -0.0505     0.0259   -1.95   0.0561 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.67 on 58 degrees of freedom
Multiple R-squared: 0.63,       Adjusted R-squared: 0.598
F-statistic: 19.8 on 5 and 58 DF,  p-value: 1.91e-11
```